doi:10.3772/j.issn.1006-6748.2025.01.001

Non-intrusive anomaly detection for carving machine systems based on CAE-GMHMM under multiple working conditions①

QIU Xiang*(仇 翔), CHEN Wei*, WU Qi②*, HU Fo*, LU Kangdi**
(*College of Information Engineering, Zhejiang University of Technology, Hangzhou 310014, P.R. China)
(**College of Information Sciences and Technology, Donghua University, Shanghai 201620, P.R. China)

Abstract

This paper is concerned with a non-intrusive anomaly detection method for carving machine systems with variant working conditions, and a novel unsupervised detection framework that integrates convolutional autoencoder (CAE) and Gaussian mixture hidden Markov model (GMHMM) is proposed. Firstly, the built-in sensor information under normal conditions is recorded, and a 1D convolutional autoencoder is employed to compress high-dimensional time series, thereby transforming the anomaly detection problem in high-dimensional space into a density estimation problem in a latent low-dimensional space. Then, two separate estimation networks are utilized to predict the mixture memberships and state transition probabilities for each sample, enabling GMHMM to handle low-dimensional representations and multi-condition information. Furthermore, a cost function comprising CAE reconstruction and GMHMM probability assessment is constructed for the low-dimensional representation generation and subsequent density estimation in an end-to-end fashion, and the joint optimization effectively enhances the anomaly detection performance. Finally, experiments are carried out on a self-developed multi-axis carving machine platform to validate the effectiveness and superiority of the proposed method.

Key words: non-intrusive detection, variant working condition, rotating machinery, motion control system, hidden Markov model (HMM)

0 Introduction

As one of the typical applications of motion control systems, multi-axis carving machine systems (MACMS) consisting of motors, sensors, drives, and transmission mechanisms are used to achieve precise position, speed, and mechanical attitude control[1]. However, various security threats are faced by networked MACMS manufacturing and processing in the industrial Internet scenario. The MACMS is not only threatened by traditional functional safety issues[2], such as equipment faults and failures, but also threatened by cyber security issues[3] like sinusoidal attacks. These issues will affect processing accuracy and product quality. Moreover, multiple working conditions may lead to different potential process characteristic modes in the manufacturing process, such as carving different materials and cutting different shapes. Therefore, it is necessary to study a new unsupervised anomaly detection method that is suitable for multiple working conditions.

Some fruitful progress has been made in the field of unsupervised anomaly detection under multiple working conditions in recent years. Cao et al.[4] utilized the Gaussian mixture model (GMM) to capture global multimodal information. However, the process data of each normal condition are supposed to follow a multivariate Gaussian distribution in GMM-based anomaly detection[5], which is often difficult to satisfy in practical scenarios. On this basis, Deng et al.[6] used the hidden Markov model (HMM) to solve the non-intrusive load monitoring problem, where the hidden state sequences are regarded as the working states of the power-using equipment. The phased results based on the HMM method have been achieved in unsupervised anomaly detection under multi-condition scenarios. However, these methods are difficult to be applied to high-dimensional data directly due to the data dimensionality and the computational complexity.

To alleviate the curse of dimensionality, two-step approaches[7-8] are widely used to generate reduced-

① Supported by the National Natural Science Foundation of China (No. 62203390).
② To whom correspondence should be addressed. E-mail: qwu@zjut.edu.cn.
Received on Mar. 25, 2024

dimensional features and then perform density estimation in a low-dimensional space. But these methods still have some drawbacks. For example, the models are decoupled, the optimization objective functions are inconsistent, and key information may be lost when it is converted to the potential low-dimensional space. Zong et al.[9] proposed an unsupervised anomaly detection method based on deep autoencoding Gaussian mixture model (DAGMM). A joint optimization strategy was adopted by DAGMM, which combined the dimensionality reduction module composed of deep autoencoder (DAE) and the density estimation module composed of GMM by a global objective function. Since the joint optimization facilitated the information transfer between the dimensionality reduction module and the density estimation module, the performance of anomaly detection was enhanced. Recently, DAGMM-based variants have emerged[10-13], mainly focusing on the improvement of the dimensionality reduction module. However, the subsequent estimation module still retains the basic GMM structure. The time-series correlation is not sufficiently considered in the DAE-based dimension compression module when processing sensor monitoring data, and the density estimation module is also deficient in variant working conditions.

On the other hand, the easiness of data acquisition is also one of the important factors to be considered. Data acquisition methods roughly consist of external sensor-based and built-in sensor-based, especially the built-in sensor-based is regarded as a non-invasive method. Yang et al.[14] proposed that deploying sensors on the drive mechanism would increase the complexity of the device. The integration and compatibility of sensors with other components need to be considered[15]. Besides, these external sensors are sensitive to environmental noise, temperature, and humidity variations, which may affect their performance and accuracy. In contrast, Huang et al.[3] addressed the problem of intrusion detection under false data injection attacks with a non-intrusive approach. This approach neither affected the functionality of the existing system nor caused security issues[16].

Motivated by the above analysis, a convolutional autoencoder (CAE)-GMHMM-based non-intrusive anomaly detection method for MACMS is proposed in this paper. Firstly, a deep one-dimensional convolutional neural network (1D-CNN)[17-19] is employed to compress the high-dimensional monitoring time series from the built-in sensors. On this basis, the problem of anomaly detection in high-dimensional space is transformed into a density estimation problem in a potential low-dimensional space. To mitigate the impact of complex variable working conditions on anomaly detection accuracy, the low-dimensional representations and temporal information are further modeled by GMHMM, facilitating the global circulation of temporal information within the model. Finally, the dimension compression module and density estimation module are synchronously optimized through an end-to-end strategy. The main contributions of this paper are summarized as follows.

(1) To address the challenge of temporal series modeling in high-dimensional space, a deep 1D convolutional autoencoder is proposed to compress the high-dimensional monitoring time series of the built-in sensors into nonlinear low-dimensional representations.

(2) For high-accuracy anomaly detection tasks under variant working conditions, the multi-axis carving process is described as a type of Markov process with unknown working condition states, while the observation sequences are only related to the working condition states.

(3) To obtain the optimal combination of compression network and estimation network, a joint optimization objective function that consists of reconstruction error, likelihood, and penalty term is designed. Besides, the network is trained by the end-to-end strategy, avoiding the need for pre-training.

1 Problem statement

1.1 Composition of research object

As shown in Fig. 1, the networked multi-axis carving machine system is mainly composed of four parts: personal computer (PC), servo driver, embedded board, and mechanical device. The PC control system is equipped with 64-bit Windows 10 operating system and necessary software simulation platforms such as Simulink, PyQt5. PC is mainly employed to realize the design and development of algorithmic simulation block diagrams and the display of the human-machine interface (HMI).

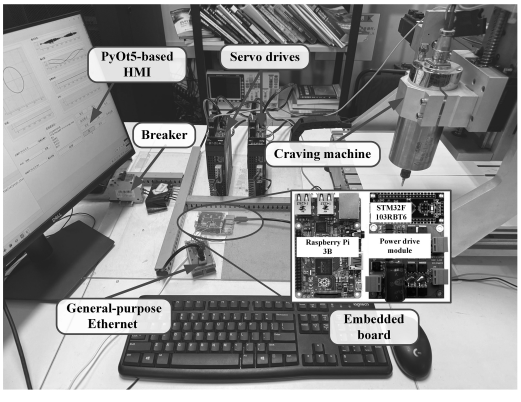

Fig. 1 Multi-axis carving machine experimental device

As a data transfer station, the embedded board, on the one hand, interacts with the servo system through its peripheral interface, receives the position, velocity, and torque information collected by the built-in sensors, and uploads them to the PC. On the other hand, it exchanges data with the PC and sends the actual control values calculated by the PC to the servo driver. The communication between the embedded board and the servo system adopts the CANopen protocol, and the data interaction between the PC and the embedded board utilizes the transmission control protocol/Internet protocol (TCP/IP).

1.2 Description of the research problem

The carving process is a discrete assembly line operation, often requiring adjustments of new equipment parameters after processing a batch of products of the same model before processing products of other models. The final product processing involves techniques such as welding and assembly. A piece of equipment often experiences different working conditions due to varying manufacturing processes. The signals of the built-in sensors under different working conditions continue to change over time, showing clear non-stationary and non-linear characteristics.

Based on Ref. [20], the anomaly occurring in MACMS is represented as

$$\boldsymbol{y}(t) = \begin{bmatrix} y_P(t) \\ y_V(t) \\ y_T(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} f(t) \tag{1}$$

where t is the time variable; $\boldsymbol{y}(t)$ is the output of the built-in sensors; $y_P(t)$, $y_V(t)$, and $y_T(t)$ are the position, velocity, and torque components of the servo motor, respectively; $f(t)$ represents the abnormal portion, and is assumed to only occur in the position sensor channel. Due to the assumption that the entire closed-loop control system is stable, anomalies in the position sensor channel are able to be detected by collecting and analyzing all built-in sensor information of each axis.

Here, the anomalies can be mainly classified into four categories including drift deviation anomaly (DDA), fixed bias anomaly (FBA), accuracy degradation anomaly (ADA), and periodic deviation anomaly (PDA)[21-22].

(1) DDA refers to a type of anomaly where the difference between the position sensor measurement and the true value changes linearly over time. For example, the harsh working environment with temperature changes is the main cause of DDA. The expression of DDA is given by

$$f(t) = w_{\text{dda}} \cdot t \tag{2}$$

where w_{dda} represents the rate of change in DDA.

(2) FBA indicates that the difference between the measurement of the position sensor and the actual value is equal to a constant. Bias current or bias voltage may be the main cause of FBA. The expression of FBA is as

$$f(t) = w_{\text{fba}} \tag{3}$$

where w_{fba} is a non-zero constant.

(3) ADA is a decline in measurement capability and precision. The average value of the position sensor measurement remains constant, but the variance increases as the accuracy grade diminishes. Electromagnetic interference, material fatigue, and sensor aging are the main factors leading to ADA. The expression of ADA is as

$$f(t) \sim w_{\text{ada}} \cdot N(0, \sigma^2) \tag{4}$$

where w_{ada} is a coefficient, $f(t)$ represents a Gaussian distribution with mean of 0 and variance of σ^2.

(4) PDA has periodicity over time, which may be caused by sinusoidal attacks. The expression of PDA is as follows.

$$f(t) = w_{\text{pda}} \cdot \sin(wt + \varphi) \tag{5}$$

where w_{pda}, w, and φ represent the amplitude, angular frequency, and phase difference, respectively.

In a closed-loop control system like MACMS, the occurrence of any anomaly will ultimately impact the system output. Sensors play a critical role in the control system as they are responsible for detecting relevant information. Therefore, the research objective of this paper is to implement non-intrusive anomaly detection in MACMS by analyzing the system output from multi-source built-in sensors.

2 Algorithm design

As shown in Fig. 2, the entire architecture includes four parts: data acquisition and processing, compression network construction, estimation network construction, and optimization objective function design. First, multi-source sensor measurements under normal conditions are obtained, including information on the position, velocity, and torque of the two axes. Next, a 1D convolutional autoencoder is utilized to generate the low-dimensional representations of the time series samples, while obtaining the reconstruction error in Euclidean space. Then, the features of multiple working condition samples in low-dimensional space are further described as a class of Markov processes with unknown working condition states. In addition, Gaussian mixture model is embedded in Markov processes to handle density estimation tasks. Finally, a global optimization objective function is designed to update the parameters of the compression and the estimation network in an end-to-end fashion.

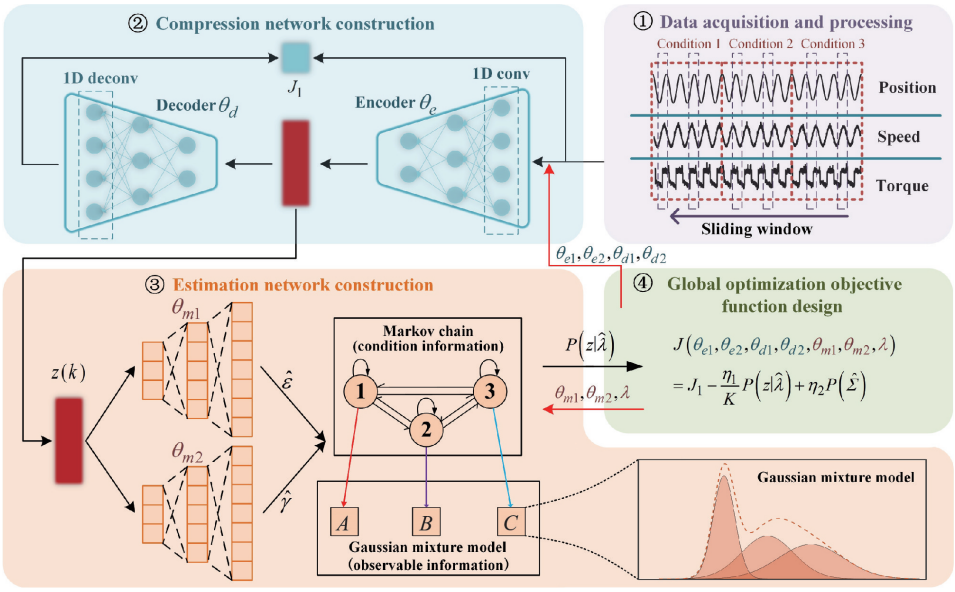

Fig. 2 Overall structure of CAE-GMHMM

Considering that there is a certain continuity in the actual carving process, the sliding window technique is performed on the multivariate built-in sensor data $\boldsymbol{y}(t) \in \mathbb{R}^{T\times D_s}$ with the window length L, where $\mathbb{R}$ represents the set of real numbers, T is the length in the time dimension, D_s is the sensor dimension. The step size is set as S, and the data after the sliding window is $\boldsymbol{y}(k,l) \in \mathbb{R}^{K\times L\times D_s}$, where $K = (T - L)/S + 1$. The working mechanism of the remaining three modules is described in the following section.

2.1 Compression network

The encoding and decoding processes are converted from regular linear layers to convolutional and deconvolutional layers by CAE for better access to the deep nonlinear temporal information. CAE consists of an encoder and a decoder. The encoder receives in-window data as input and generates new low-dimensional representations after encoding in 1D convolution. The encoding process is represented as

$$z_e = MLN(1\text{D } conv(y,\theta_{e1}),\theta_{e2}) \tag{6}$$

where 1D $conv(\cdot)$ denotes 1D convolution and $MLN(\cdot)$ is fully connected layer; θ_{e1} and θ_{e2} are the parameters of convolution layer and fully connected layer, respectively; $z_e \in \mathbb{R}^N$ is the low-dimensional representations learned by the deep autoencoder with dimensions N. Since the multivariate time series data contains multiple channels $D = D_s$, multi-channel kernels are created to process the input data.

The decoder reconstructs the original input data through the low-dimensional representations obtained from the encoder after 1D deconvolution, and the decoding process is represented as

$$y_d = 1\text{D } deconv(MLN(z_e,\theta_{d1}^{\mathrm{T}}),\theta_{d2}^{\mathrm{T}}) \tag{7}$$

where 1D $deconv(\cdot)$ denotes 1D deconvolution; θ_{d1}^{T} and θ_{d2}^{T} are the parameters of the fully connected layer and deconvolution layer, respectively; y_d is the reconstructed sample data.

CAE dimension reduction process is shown in Fig. 3. The training objective of CAE is to minimize the reconstruction error, and the mean square error is used as the loss function, which is expressed as

$$J_1 = \frac{1}{K}\times\frac{1}{L}\sum_{k=1}^{K}\sum_{l=1}^{L}\| y(k,l),y_d(k,l)\|_2^2 \tag{8}$$

where $y(k,l)$ denotes the input samples after sliding window at moment k of window j, and $y_d(k,l)$ denotes the reconstructed samples, $\|\cdot\|_2^2$ represents the L2 norm.

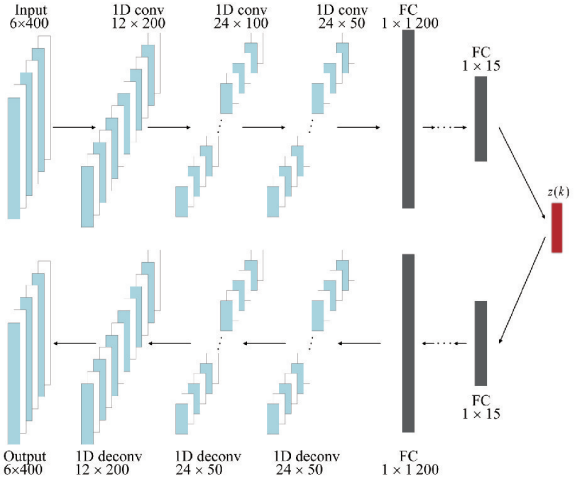

Fig. 3 Multi-source sensor measurements dimension reduction process based on 1D convolutional autoencoder

2.2 Estimation network

The original DAGMM anomaly detection method utilizes GMM as the estimation network to obtain the distribution of signal energy values, which enhances the discrimination of some anomaly samples. However, in order to enable the model to deal with multi-condition information, GMHMM is adopted to capture the time dependence and correlation between variables in the time series to fully utilize the hidden information.

Similarly, the estimation network is employed to estimate the GMHMM parameters without the alternating algorithm of the expectation maximization.

First, given the model and observation sequences, two multi-layer neural networks are utilized to predict the state transition probabilities $\hat{\boldsymbol{\varepsilon}}$ and the mixture memberships $\hat{\boldsymbol{\gamma}}$. The mathematical expression of estimation network is as

$$\hat{\boldsymbol{\varepsilon}} = \mathrm{softmax}(MLN(z_e, \theta_{m1})) \tag{9}$$

$$\hat{\boldsymbol{\gamma}} = \mathrm{softmax}(MLN(z_e, \theta_{m2})) \tag{10}$$

where θ_{m1} and θ_{m2} are the parameters of the two multi-layer networks. $\hat{\boldsymbol{\varepsilon}}(k)$ represents the state transition matrix of $U \times U$ dimensions, and $\hat{\boldsymbol{\gamma}}(k)$ represents the membership matrix of $U \times M$ dimensions; U is the number of hidden states, and M is the number of Gaussian distributions.

Next, $\hat{\lambda} = (\hat{\pi}, \hat{a}, \hat{c}, \hat{\mu}, \hat{\boldsymbol{\Sigma}})$ is estimated by $\hat{\boldsymbol{\varepsilon}}$ and $\hat{\boldsymbol{\gamma}}$ obtained from the estimation network, and the GMM is used to fit the probability density function of the observations in each hidden state.

$$\hat{\pi}_i = \sum_{m=1}^{M} \hat{\gamma}_{i,m}(1), \hat{a}_{i,j} = \frac{\sum_{k=1}^{K-1} \hat{\varepsilon}_{i,j}(k)}{\sum_{k=1}^{K-1}\sum_{m=1}^{M} \hat{\gamma}_{i,m}(k)} \tag{11}$$

$$\hat{c}_{i,m} = \frac{\sum_{k=1}^{K} \hat{\gamma}_{i,m}(k)}{\sum_{k=1}^{K}\sum_{m=1}^{M} \hat{\gamma}_{i,m}(k)}, \hat{\mu}_{i,m} = \frac{\sum_{k=1}^{K} \hat{\gamma}_{i,m}(k) \cdot z(k)}{\sum_{k=1}^{K} \hat{\gamma}_{i,m}(k)} \tag{12}$$

$$\hat{\boldsymbol{\Sigma}}_{i,m} = \frac{\sum_{k=1}^{K} \hat{\gamma}_{i,m}(k)(z(k) - \hat{\mu}_{i,m})(z(k) - \hat{\mu}_{i,m})^{\mathrm{T}}}{\sum_{k=1}^{K} \hat{\gamma}_{i,m}(k)} \tag{13}$$

where, $\hat{\pi}_i$, $\hat{a}_{i,j}$, $\hat{c}_{i,m}$, $\hat{\mu}_{i,m}$ and $\hat{\boldsymbol{\Sigma}}_{i,m}$ denote the initial state, state transfer probability, mixture probability, mean, and covariance, respectively; $i = 1, 2, \cdots, U$; $j = 1, 2, \cdots, U$; $m = 1, 2, \cdots, M$.

Then the probability $\boldsymbol{B} = [b_i(z(k))]_{U \times K}$ is calculated according to Eqs. (12) and (13).

$$b_i(z(k)) = \sum_{m=1}^{M} \hat{c}_{i,m} \frac{\exp\left(-\frac{1}{2}(z(k) - \hat{\mu}_{i,m})^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_{i,m}^{-1}(z(k) - \hat{\mu}_{i,m})\right)}{\sqrt{|2\pi\hat{\boldsymbol{\Sigma}}_{i,m}|}} \tag{14}$$

where $|\cdot|$ is the determinant of the matrix. Then, two variables including the forward probability $\alpha_i(k)$ and the backward probability $\beta_i(k)$ are defined and computed. The forward probability is defined as

$$\alpha_i(k) = P(z(1), z(2), \cdots, z(k), q(k) = s_i \mid \hat{\lambda}) \tag{15}$$

where $q(k)$ represents the hidden state at moment k, s_i is the ith hidden state. The recursive formula is $\alpha_i(k) = \sum_{j=1}^{U} \alpha_j(k-1)\hat{a}_{j,i} b_i(z(k))$, and the initial value is $\alpha_i(1) = \hat{\pi}_i b_i(z(1))$. The backward probability is

$$\beta_i(k) = P(z(k+1), z(k+2), \cdots, z(K) \mid q(k) = s_i, \hat{\lambda}) \tag{16}$$

The recursion is $\beta_i(k) = \sum_{j=1}^{U} \hat{a}_{j,i} b_j(z(k+1)) \times \beta_j(k+1)$, and initial value of $\beta_i(K) = 1$.

Finally, the likelihood is computed based on $\alpha_i(k)$ and $\beta_i(k)$ [23].

$$P(z \mid \hat{\lambda}) = \sum_{i=1}^{U} \alpha_i(k)\beta_i(k) \tag{17}$$

where $P(z \mid \hat{\lambda})$ is the likelihood probability value. For normal samples, a large likelihood can be obtained by entering them into GMHMM. In contrast, abnormal samples result in a lower likelihood, and the likelihood tends to decrease as the magnitude of the abnormality increases.

2.3 Objective function

So far, the two-step method of CAE downgrading and GMHMM estimation has been completed, but they are carried out independently. The CAE downgrading process is trained without the guidance of GMHMM, and the key information may be lost during dimensionality reduction. To address this issue, the global optimization objective function is designed to induce representations of potentially low-dimensional features that are utilized for subsequent estimation tasks. The global objective function is constructed as follows.

$$\begin{aligned} &J(\theta_{e1}, \theta_{e2}, \theta_{d1}, \theta_{d2}, \theta_{m1}, \theta_{m2}, \lambda) \\ &= J_1 - \frac{\eta_1}{K} P(z \mid \hat{\lambda}) + \eta_2 P(\hat{\Sigma}) \\ &= \frac{1}{K} \times \frac{1}{L} \sum_{k=1}^{K} \sum_{l=1}^{L} \| y(k,l) - y_d(k,l) \|_2^2 - \\ &\quad \frac{\eta_1}{K} \sum_{i=1}^{U} \alpha_i(k)\beta_i(k) + \eta_2 \sum_{i=1}^{U} \sum_{m=1}^{M} \sum_{n=1}^{N} \frac{1}{\Delta_{n,n}} \end{aligned} \tag{18}$$

where θ_{d1} and θ_{d2} are the CAE parameters. The representation of $\hat{\boldsymbol{\Sigma}}_{i,m}$ is

$$\hat{\boldsymbol{\Sigma}}_{i,m} = \begin{bmatrix} \Delta_{1,1} & & \\ & \ddots & \\ & & \Delta_{N,N} \end{bmatrix}_{N\times N} \tag{19}$$

where $\Delta_{n,n}$ denote the diagonal values of $\hat{\boldsymbol{\Sigma}}_{i,m}$, N is the dimension of the low-dimensional representations.

The optimization objective function consists of three parts: reconstruction error, likelihood, and penalty term. $\| y(k,l) - y_d(k,l) \|_2^2$ is a distance measurement between the input samples and the reconstructed samples, which is used to characterize the reconstruction error introduced by the 1D convolutional autoencoder. $P(z|\hat{\lambda})$ is computed from $\alpha_i(k)$ and $\beta_i(k)$, representing the probability of the observations when the model parameters are known. By maximizing $P(z|\hat{\lambda})$, it is promising to find the optimal combination of the compression and estimation networks that maximizes the likelihood of observing input samples. In addition, to avoid trivial solutions when computing the parameters of the GMHMM, a penalty term $P(\hat{\Sigma})$ is utilized to penalize small values on diagonal entries.

Remark 1 The optimization objective function with similar structure is common in the Refs. [9-13]. Firstly, ignoring the reconstruction error will result in the compression network not effectively capturing the key features of the samples, which in turn affects the quality of low-dimensional representations. Furthermore, likelihood is one of the prerequisites to ensure that the model is able to handle the anomaly detection task for multiple conditions. Finally, the penalty term is designed to avoid the singularity problem in the covariance matrix, thus ensuring the stability and reliability of the model.

The non-invasive anomaly detection process based on CAE-GMHMM is summarized as Algorithm 1, which is specifically divided into offline training and online detection.

(1) Offline training phase: the objective function hyperparameters η_1 and η_2 are predefined, and the network parameters θ_{e1}, θ_{e2}, θ_{d1}, θ_{d2}, θ_{m1}, and θ_{m2} are initialized. Then, the network parameters and GMHMM parameters $\hat{\pi}_i$, $\hat{a}_{i,j}$, $\hat{c}_{i,m}$, $\hat{\mu}_{i,m}$, and $\hat{\Sigma}_{i,m}$ are continuously updated according to the training set data to complete the training. The ζth percentile of the output likelihood of all training samples is set as the anomaly detection threshold.

Anomaly detection process based on CAE-GMHMM is shown in Algorithm 1.

Algorithm 1 Anomaly detection process based on CAE-GMHMM

Input: Multi-source sensor dataset $y(l)$, initial network parameters including θ_{e1}, θ_{e2}, θ_{d1}, θ_{d2}, θ_{m1}, and θ_{m2}, hyperparameters such as η_1 and η_2.

Output: Detected result.

Offline training

1. Generate the low-dimensional representations z_e and reconstruction error J_1 by Eqs. (6) - (8);
2. Predict the probabilities $\hat{\boldsymbol{\varepsilon}}$ of state transfer and the mixture memberships $\hat{\boldsymbol{\gamma}}$ from Eqs. (9) and (10);
3. Obtain maximum likelihood estimation of GMHMM parameters $\hat{\pi}_i$, $\hat{a}_{i,j}$, $\hat{c}_{i,m}$, $\hat{\mu}_{i,m}$, and $\hat{\Sigma}_{i,m}$ from Eqs. (11) - (13);
4. Compute the forward probability $\alpha_i(k)$ and backward probability $\beta_i(k)$ from Eq. (14) and Eq. (16);
5. Calculate the likelihood $P(z|\hat{\lambda})$ via Eq. (17);
6. Compute global loss $J(\theta_{e1},\theta_{e2},\theta_{d1},\theta_{d2},\theta_{m1},\theta_{m2},\lambda)$ based on J_1, $P(z|\hat{\lambda})$, and $P(\hat{\Sigma})$ via Eq. (18);
7. Update network parameters θ_{e1}, θ_{e2}, θ_{d1}, θ_{d2}, θ_{m1}, and θ_{m2} with Adam optimizer;
8. Repeat the above steps until the convergence of the algorithm is reached;
9. Obtain the optimal compression parameters θ_{e1}, θ_{e2}, GMHMM parameters $\hat{\pi}_i$, $\hat{a}_{i,j}$, $\hat{c}_{i,m}$, $\hat{\mu}_{i,m}$, $\hat{\Sigma}_{i,m}$, and threshold v.

Online detection

10. Extract low-dimensional representations z_e via Eq. (6);
11. Calculate the likelihood $P(z|\hat{\lambda})$ from Eqs. (14) - (17) and parameters $\hat{\pi}_i$, $\hat{a}_{i,j}$, $\hat{c}_{i,m}$, $\hat{\mu}_{i,m}$, and $\hat{\Sigma}_{i,m}$;
12. If $P(z|\hat{\lambda}) < v$ then
Result = 1 (Abnormal);
13. Else
Result = 0 (Normal).

(2) Online detection phase: the low-dimensional representations of the test samples are obtained through the compression network. Then, the likelihood of the test samples is calculated according to the low-dimensional representations and the trained GMHMM parameters. Finally, compared with the threshold of all training sample likelihood, samples below the threshold are considered anomalous.

3 Case study

3.1 Case data

A self-developed multi-axis carving machine system is used to validate the effectiveness and superiority of the proposed CAE-GMHMM method. To simulate the multiple working condition carving process in the physical world, various circular trajectory machining tasks

with different parameters are carried out. In the experiment, circles with amplitudes of 10.0 cm, 9.5 cm, and 9.0 cm are sequentially processed. The circular trajectory processing cycle of each parameter is set to 100 times.

Taking the circle with the amplitude of 10.0 cm as an example, the tracking trajectory can be defined as

$$\begin{cases} ref_x = 10\sin(2\pi t/1600) + 20 \\ ref_y = 10\cos(2\pi t/1600) + 20 \end{cases} \tag{20}$$

where the center of the circle is (20,20). Then the circle tracking task containing abnormal samples is carried out with the parameter configurations shown in Table 1.

Table 1 Abnormal parameter configurations

Abnormal type	Abnormal description	Abnormal manifestation
NS	Normal state	No abnormality occurs
DDA	Drift deviation anomaly	$f(t) = 0.002\,5t$
FBA	Fixed bias anomaly	$f(t) = 0.125$
ADA	Accuracy degradation anomaly	$f(t) \sim 0.005N(0,1)$
PDA	Periodic deviation anomaly	$f(t) = 0.001\sin(2\pi/200 \times t)$

Part of built-in sensor measurements under four kinds of anomalies are shown in Fig. 4. The light-shaded areas indicate that the system is under abnormal conditions, while the white areas indicate that the system is under normal conditions. Different types of anomalies are simulated and injected into the position sensor channel at the 4 800th sampling point. Drift deviation anomaly is shown in Fig. 4(a). Such type of anomaly is directly proportional to time. The longer it lasts, the more severe the impact on the equipment becomes, potentially leading to sudden changes in motor speed. Fixed bias anomaly is illustrated in Fig. 4(b), and it is observed that the motor speed suddenly changes the first time the anomaly takes effect. The degradation of the sensor measurement capability and accuracy is shown in Fig. 4(c). It is found that the sensor introduces noise with large variance. In Fig. 4(d), the sensor measurement with periodic deviation anomaly is displayed, and periodic fluctuations in the data are observed. In a closed-loop control system, the system still achieves good trajectory tracking performance due to the robustness of the control algorithm. However, prolonged exposure to abnormal conditions will accelerate equipment fatigue and wear.

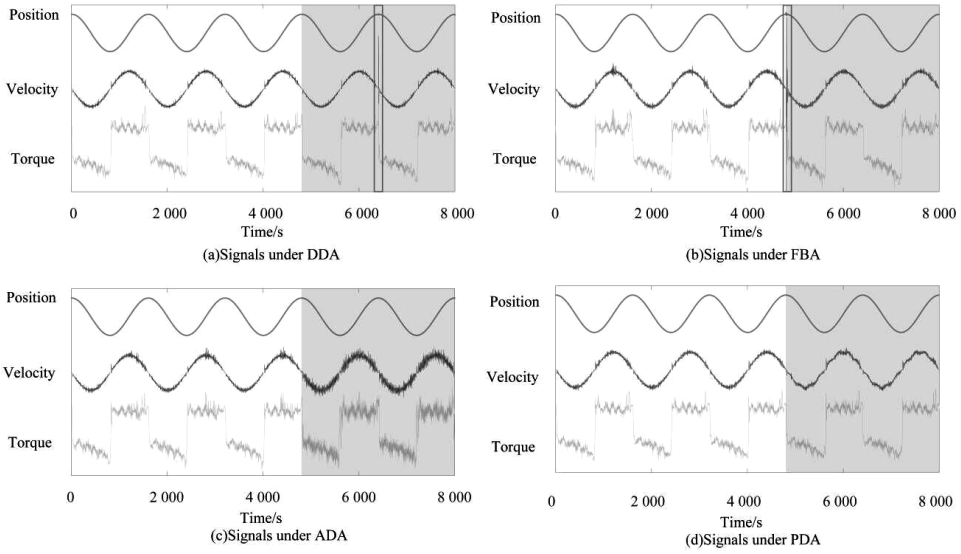

Fig. 4 Built-in sensor measurements under four kinds of anomalies

The purpose of this paper is to analyze the data obtained from built-in sensor measurements for detection. The position, velocity, and torque data of x axis and y axis are collected in real-time from the built-in sensors of MACMS for the experiments with a sampling interval of 5 ms. In the data preprocessing process, the data of each type of built-in sensor signals are intercepted without overlapping parts using sliding windows, and the length of each sliding window is $L = 400$. Taking the first working condition as an example, anomalies occurred after carving 100 circles of normal state (NS), and each anomaly lasted for 400 s. Therefore, 1 200 NS samples and 300 abnormal samples of each kind were intercepted under three working conditions. 50% of the normal samples are randomly selected for training, and the remaining 50% are reserved for testing, using on-

ly normal samples for model training. All abnormal samples and 50% normal samples are taken for model testing.

According to the mainstream anomaly detection evaluation protocols[24-26], the anomaly detectors in the experiments are evaluated by the area under the ROC curve (AUC-ROC), the area under the PR curve (AUC-PR), and $F1$ score. The AUC-ROC summarizes the ROC curves for true positives and false positives, AUC-PR is a performance metric that focuses more on anomalies and summarizes the curves for accuracy and recall, and the $F1$ score combines accuracy and recall, which is a balance between them.

3.2 Comparison test

In this section, the proposed method is compared with other unsupervised deep anomaly detection methods including deep support vector data description (DSVDD)[27], robust collaborative autoencoders (RCA)[28], customized representations for a random nearest neighbor distance-based method (REPEN)[29], deep isolation forest (DIF)[30], DAGMM, CAE-GMM, and autoencoder (AE)-GMHMM. DSVDD is an end-to-end version of support vector data description (SVDD) for deep modeling that minimizes the hypersphere volume of the network representation during neural network training. RCA trains a set of autoencoders in a collaborative manner and learns their model parameters and sample weights together. Given a small batch, each autoencoder learns feature representations and selects the subset of samples with the lowest reconstruction error. It selects samples from each autoencoder and exchanges samples among them to update their model weights to avoid premature convergence. REPEN implements random distance-based outlier detection by learning low-dimensional representations of high-dimensional data, unifying the two related tasks of representation learning and anomaly detection. Instead of describing the difference between samples with the help of metrics such as distance or density, DIF utilizes the powerful representation ability of neural networks to map the original data into a set of new spaces. It judges the anomalies by carving out their sparsity on these new spaces to achieve nonlinear segregation. DAGMM is the basic framework, which balances autoencoding reconstruction, density estimation, and regularization. CAE-GMM is an improved DAGMM that encodes and decodes the time-series features of multivariate time series by 1D convolution during feature extraction. AE-GMHMM utilizes two estimation networks and the probabilistic model is replaced with GMHMM.

The network structure of CAE consists of 3 convolutional layers, 8 fully connected layers (FC), and 3 deconvolutional layers. The numbers of convolutional kernels are 12, 24, and 24 respectively. The size of convolutional kernel is 3, the step size is 1, and the padding is 1. The nonlinear activation function is chosen as ReLU used in all convolutional neural network(CNN) layers and max-pooling is applied for downsampling. Dropout is used to prevent over-adjustment on training data. Eight fully connected layers run with FC(1 200, 150) – FC(150, 100) – FC(100, 50) – FC(50, 15) – FC(15, 50) – FC(50, 100) – FC(100, 150) – FC(150, 1 200). Estimation network structures consist of two fully connected networks. $\hat{\boldsymbol{\varepsilon}}(k)$ performs with FC(15, 20) – FC(20, 100) and $\hat{\boldsymbol{\gamma}}(k)$ runs with FC(15, 20) – FC(20, 100). The Adam optimizer is used to optimize the parameters of the model with a learning rate of 0.001. The number of hidden states is $U = 10$, and the number of Gaussian distributions is $M = 10$. The number of low-dimensional features is $N = 15$. The hyperparameters of the optimization objective function are $\eta_1 = 1 \times 10^{-7}$ and $\eta_2 = 1 \times 10^{-9}$ respectively. The $\zeta = 5$ percentile of all training sample likelihood is utilized as the anomaly detection threshold. The framework used for deep learning is Pytorch 1.12. The computer configuration for executing the program is as follows: Intel Core i7-12700, NVIDIA RTX3080, 16 GB RAM.

The results of AUC-ROC, AUC-PR, and $F1$ scores for the compared methods and the proposed CAE-GMHMM are shown in Table 2. As expected, the best performance is achieved by the CAE-GMHMM in most anomalous cases. DSVDD performs moderately because it is difficult to find the boundary between abnormal and normal samples. For REPEN, it introduces the outlier detection based on random distance into its objective function, but when faced with time series data, it may not accurately capture the relationship between features. Although RCA learned the hidden representations in data, it fails to consider the time series correlation. DIF shows strong performance in detection, but its generalization ability is poor when faced with complex working conditions. In DAGMM, the temporal correlation and the effect of multiple working conditions are not considered. CAE-GMM cannot effectively reflect the multi-working conditions. For AE-GMHMM, the temporal correlation is disregarded during feature compression, leading to potentially small reconstruction error for anomalies. These last four methods also constitute an ablation experiment, further validating the effectiveness and superiority of the proposed method.

Table 2 Comparison of detection performance between CAE-GMHMM and other methods

	DDA			FBA			ADA			PDA		
	AUC-RUC	AUC-PR	*F*1	AUC-RUC	AUC-PR	*F*1	AUC-RUC	AUC-PR	*F*1	AUC-RUC	AUC-PR	*F*1
DSVDD	0.738 9	0.853 6	0.747 5	0.759 5	0.843 3	0.797 5	0.814 4	0.894 8	0.807 5	0.970 6	0.986 8	0.932 5
RCA	0.565 5	0.725 5	0.697 5	0.507 8	0.693 8	0.672 5	0.898 5	0.950 3	0.855 0	0.915 3	0.953 4	0.877 5
REPEN	0.531 9	0.676 4	0.602 5	0.572 0	0.636 7	0.655 0	0.778 1	0.884 1	0.757 5	0.909 7	0.959 1	0.842 5
DIF	0.735 8	0.853 1	0.732 5	0.926 3	0.939 4	0.922 5	0.725 5	0.873 1	0.827 5	0.974 9	0.982 8	0.960 0
DAGMM	0.537 5	0.606 2	0.558 3	0.527 5	0.679 2	0.685 0	0.567 5	0.698 7	0.711 7	0.616 2	0.724 3	0.744 2
AE-GMHMM	0.751 7	0.801 1	0.859 5	0.866 7	0.891 5	0.903 8	0.906 3	0.919 2	0.923 7	0.893 7	0.915 4	0.940 6
CAE-GMM	0.747 1	0.813 8	0.751 9	0.753 8	0.827 3	0.831 4	0.852 5	0.881 4	0.893 0	0.821 2	0.875 9	0.911 8
CAE-GMHMM	0.976 0	0.987 9	0.945 0	0.955 4	0.978 1	0.952 5	0.989 0	0.995 0	0.982 5	0.998 8	0.999 3	0.997 5

3.3 Deep feature visualization results

In order to intuitively visualize the deep features obtained by compression, the t-distributed stochastic neighbor embedding (t-SNE)[31] is employed to present the learned feature distribution.

Taking the detection of FBA as an example, the feature visualization results of the four algorithms are shown in Fig. 5, respectively. The implied feature distributions of normal samples under the three working conditions of DAGMM are scattered, and there is some overlap between normal and abnormal samples in Region 1. The feature distances of normal samples in different working conditions of AE-GMHMM are narrowed due to the consideration of the condition factor. CAE-GMM has a concentrated feature distribution with a long

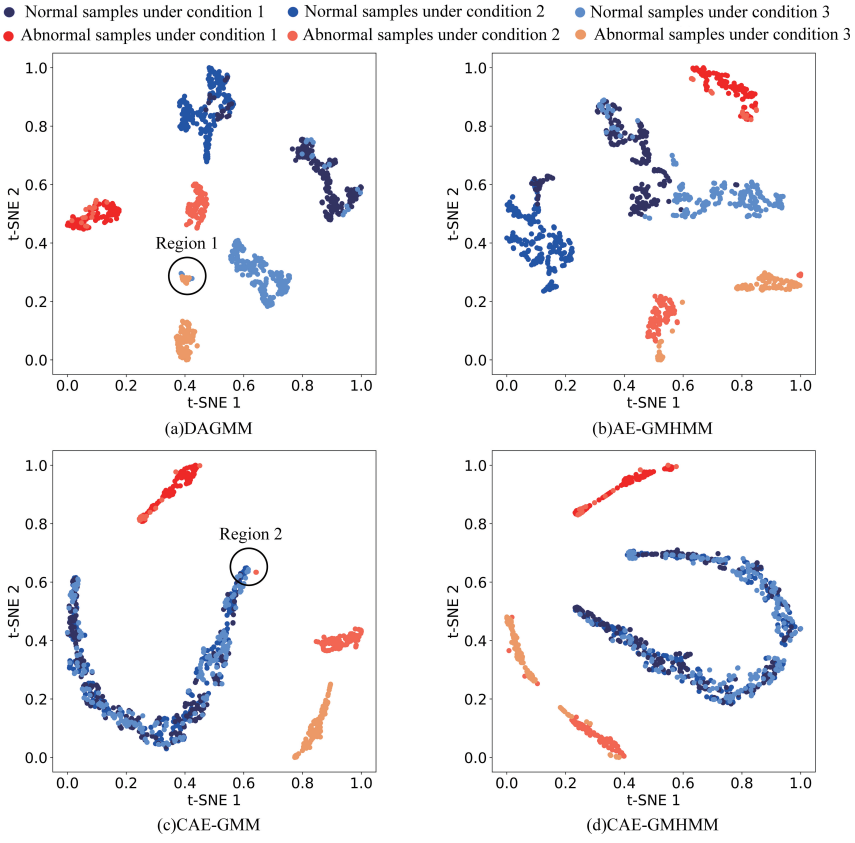

Fig. 5 t-SNE visualization results

strip on the normal samples due to the extraction of the temporal features, but there is still some overlap as shown in Region 2, and the unknown anomalous samples are still mixed in the known normal samples. Fortunately, the low-dimensional representations learned by CAE-GMHMM are separable. Considering variant working scenarios, the normal samples are basically connected with each other, the distribution is more concentrated, and the cluster contains all the normal samples, which is well recognized by GMHMM to identify the normal and abnormal samples. Therefore, the deep features learned by CAE-GMHMM are better than the deep features learned by the other three algorithms.

4 Conclusions

In this work, a non-intrusive anomaly detection method for MACMS based on CAE-GMHMM is proposed. Multivariate built-in sensor measurements are adopted as input, and 1D convolutional autoencoder and Gaussian mixture hidden Markov model for low-dimensional spatial density estimation are integrated. Specifically, 1D convolutional encoder is utilized to improve the ability to extract time series features during data dimensionality reduction. The GMHMM is employed to model the low-dimensional representation which is beneficial for the global flow of time series information within the model. Then, the parameters of CAE and GMHMM are jointly optimized in end-to-end training. Finally, through comparative experiments conducted on the experimental platform, the experiment results show that the proposed method is superior to state-of-the-art methods.

Reference

[1] WU X, WANG J X, WANG Y W, et al. ADRC-based high precision contour tracking control of networked motion control system[J]. High Technology Letters, 2018, 28(Z1): 835-842. (In Chinese)

[2] WU Q, DONG C, GUO F H, et al. Privacy-preserving federated learning for power transformer fault diagnosis with unbalanced data[J]. IEEE Transactions on Industrial Informatics, 2023, 20(4): 5383-5394.

[3] HUANG D J, SHI X F, ZHANG W A, False data injection attack detection for industrial control systems based on both time-and frequency-domain analysis of sensor data[J]. IEEE Internet of Things Journal, 2020, 8(1): 585-595.

[4] CAO Y, JAN N M, HUANG B, et al. No-delay multimodal process monitoring using Kullback-Leibler divergence-based statistics in probabilistic mixture models[J]. IEEE Transactions on Automation Science and Engineering, 2022, 20(1): 167-178.

[5] TAN S, CHANG Y Q, WANG F L, et al. Mode identification and process monitoring for multiple mode processes based on GMM[J]. Control and Decision, 2015, 30(1): 53-58.

[6] DENG X P, ZHANG G Q, WEI Q L, et al. A survey on the non-intrusive load monitoring[J]. Acta Automatica Sinica, 2022, 48(3): 644-663.

[7] CHANDOLA V, BANERJEE A, KUMAR V. Anomaly detection: a survey[C]//ACM Computing Surveys. New York, USA: ACM, 2009: 1-58.

[8] CHALAPATHY R, CHAWLA S. Deep learning for anomaly detection: a survey[EB/OL]. (2019-01-10) [2024-03-15]. https://arxiv.org/pdf/1901.03407.

[9] ZONG B, SONG Q, MIN M R, et al. Deep autoencoding Gaussian mixture model for unsupervised anomaly detection[C]// International Conference on Learning Representations. Vancouver, Canada: OpenReview, 2018: 1-19.

[10] TIAN Y, LI J L, SONG Q Z, et al. Pyramid reconstruction assisted deep autoencoding Gaussian mixture model for industrial fault detection[J]. Information Sciences, 2023, 649: 119682.

[11] CHEN Y, ZHANG J Z, YEO C K. Network anomaly detection using federated deep autoencoding Gaussian mixture model[C]//International Conference on Machine Learning for Networking. Paris, France: Springer, 2019: 1-14.

[12] PUROHIT H, TANABE R, ENDO T, et al. Deep autoencoding GMM-based unsupervised anomaly detection in acoustic signals and its hyper-parameter optimization[EB/OL]. (2020-09-25) [2024-03-15]. https:// arxiv.org/pdf/2009.12042.

[13] CHEN Y, ASHIZAWA N, YEAN S, et al. Self-organizing map assisted deep autoencoding Gaussian mixture model for intrusion detection[C]//2021 IEEE 18th Annual Consumer Communications and Networking Conference. Las Vegas, USA: IEEE, 2021: 265-270.

[14] YANG M, DONG C Y, XU D G, et al. Review of gear fault diagnosis methods based on motor drive system[J]. Transactions of China Electrotechnical Society, 2016, 31(4): 58-63.

[15] CUI C. Study on key technologies in non-intrusive residential load monitoring for intelligent power utilization[D]. Beijing: North China Electric Power University, 2017. (In Chinese)

[16] WU Z. Study on key technologies in non-intrusive load disaggregation for intelligent power utilization[D]. Chongqing: Chongqing University, 2022. (In Chinese)

[17] HANG G J, TIAN C, GOU L F, et al. Classification for aero-engine fault modes under multiple operating conditions[J]. Journal of Chinese Computer Systems, 2022, 43(8): 1776-1781. (In Chinese)

[18] XIE X, WANG B, WAN T C, et al. Multivariate abnormal detection for industrial control systems using 1D CNN and GRU[J]. IEEE Access, 2020, 8: 88348-88359.

[19] KRAVCHIK M, SHABTAI A. Detecting cyber attacks in industrial control systems using convolutional neural networks[C]//Proceedings of the 2018 Workshop on Cyber-physical Systems Security and Privacy. Toronto, Canada: ACM, 2018: 72-83.

[20] WU Q, YU L, WANG Y W, et al. LESO-based position

synchronization control for networked multi-axis servo systems with time-varying delay[J]. IEEE/CAA Journal of Automatica Sinica, 2020, 7(4): 1116-1123.
[21] JIA Z, LI Y, WANG S, et al. A multi-channel data-based fault diagnosis method integrating deep learning strategy for aircraft sensor system[J]. Measurement Science and Technology, 2022, 34(2): 025115.
[22] HE F J, LIU X L, LU X Z, et al. A research on fault detection diagnosis technology for sensor[J]. Science Technology and Engineering, 2010, 10(26): 1671-1815.
[23] BAUM L E, EAGON J A. An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology[J]. Bulletin of the American Mathematical Society, 1967, 73(3): 360-363.
[24] LIU F T, TING K M, ZHOU Z H. Isolation forest[C]// 2008 8th IEEE International Conference on Data Mining. Pisa, Italy: IEEE, 2008: 413-422.
[25] HARIRI S, KIND M C, BRUNNER R J. Extended isolation forest[J]. IEEE Transactions on Knowledge and Data Engineering. 2019, 33(4): 1479-1489.
[26] PANG G S, SHEN C H, HENGEL A V D. Deep anomaly detection with deviation networks[C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Anchorage, USA: ACM, 2019: 353-362.
[27] RUFF L, VANDERMEULEN R, GOERNITZ N, et al. Deep one-class classification[C]//International Conference on Machine Learning. Stockholm, Sweden: ACM, 2018: 4393-4402.
[28] KWON H Y, KIM T, LEE M K. Advanced intrusion detection combining signature-based and behavior-based detection methods[J]. Electronics, 2022, 11(6): 867.
[29] PANG G S, CAO L B, CHEN L, et al. Learning representations of ultrahigh-dimensional data for random distance-based outlier detection[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. London, UK: ACM, 2018: 2041-2050.
[30] XU H Z, PANG G S, WANG Y J, et al. Deep isolation forest for anomaly detection[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(12): 12591-12604.
[31] MAATEN L V D, HINTON G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(11): 2579-2605.

QIU Xiang, born in 1980. He received the Ph. D degree in control theory and engineering from Zhejiang University of Technology, Hangzhou, China, in 2016. He is currently an associate professor in the Department of Automation at Zhejiang University of Technology, Hangzhou, China. His current research interests include data analysis, intelligent control, industrial robot, complex networks.