Hindawi
Security and Communication Networks
Volume 2022, Article ID 3757662, 14 pages
https://doi.org/10.1155/2022/3757662

Research Article

Feature Extraction Method Based on Sparse Autoencoder for Air Traffic Management System Security Situation Awareness

Zhijun Wu,[1] **Zhuoning Bai**,[1] **Lizhe Zhang**,[1,2] **and Kenian Wang**[1,2]

[1]*School of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China*
[2]*Key Laboratory of Civil Aircraft Airworthiness Technology, Civil Aviation University of China, Tianjin 300300, China*

Correspondence should be addressed to Lizhe Zhang; lzzhang@cauc.edu.cn

Received 27 June 2022; Revised 26 July 2022; Accepted 11 August 2022; Published 5 September 2022

Academic Editor: Shudong Li

In wide-area distributed scenarios, it is particularly important to carry out information security situational awareness for the air traffic management (ATM) system with integrated air-ground structure. The operation data of the communication, navigation and surveillance (CNS) equipment of ATM system have the characteristics of multi-dimension, complexity, and strong correlation. In the process of situation awareness feature extraction, there are problems such as poor model accuracy, weak feature expression ability, and low classification performance. A feature association algorithm is designed to solve the above problems. Based on this algorithm, a deep-related sparse autoencoder (DRSAE) model based on improved sparse autoencoder is established. In DRSAE model, L_1 regularization and Kullback–Leibler divergence (KLD) sparsity terms are used to penalize the parameters of the encoder network, and the quantity of hidden layers is increased to allow the model to optimize the global encoder network by iteratively training a single encoder. Moreover, the proposed DRSAE model and other feature extraction models such as principal component analysis (PCA), autoencoder (AE), and sparse autoencoder (SAE) are compared and evaluated by using the support vector machine (SVM) classifier. Compared with other feature extraction models, it is found that the proposed DRSAE model has good robustness in feature extraction of ATM system, and the obtained features have strong expression ability, which enhances the classification performance of the model and is convenient for situation awareness.

1. Introduction

As the key equipment of air traffic management (ATM) system, the safe operation of the communication, navigation and surveillance (CNS) equipment is very important to ensure the stable information exchange of ATM system. CNS equipment contains various types of ATM data, and its security involves all levels of the ATM system. If the confidential information is subject to the abovementioned security threats, it will affect the flight status of the aircraft and even threaten the safety of aircraft navigation in severe cases. In addition, when the equipment fails due to human or design reasons, the ATM information exchange business will be suspended. In the process of information interactive transmission, the security threats of CNS equipment of ATM system generally include insecure physical access to IT infrastructure, unencrypted communication of satellite or ground system, insufficient security configuration of IT equipment hardware and software, unauthorized personnel access, malware infection, etc. [1].

The high-dimensional and complex operation data of ATM system equipment make it difficult for decision-making managers to perceive the security situation of the system. To make sure the availability, confidentiality, and integrity of the operation information of ATM system equipment, it is imperative to perceive the information security situation of ATM system. The situation awareness of CNS equipment aims to reduce the effect of the equipment on the operation of ATM system due to network threats, human operation errors, equipment failures, etc., so that managers can timely understand the security status of the system from a macro perspective and make correct decision [2]. In situation awareness research, feature extraction of data is the first step to obtain situation elements. Feature

extraction transforms multi-dimensional feature space into universal low-dimensional feature space through linear transformation or nonlinear transformation, which is convenient for the subsequent situation assessment of the system [3].

The ATM system involves a wide range of users, so it is the most basic to guarantee the data security in the ATM system. The paper starts with the communication system equipment, navigation system equipment, surveillance network such as Automatic Dependent Surveillance-Broadcast (ADS-B) system, and management network such as System-Wide Information Management (SWIM) system platform of ATM system, analyzes the security threats faced by each system equipment and the existing system vulnerabilities, and studies the feature extraction of equipment operation data. The main innovations are as follows:

(i) Since the operation data of ATM equipment have the characteristics of multi-dimensional, numerous, time-space dependence, etc., and there is correlation between the features, in order to prevent overfitting and high complexity, the data should be dimensionally reduced to use representable simple dimensional structure to represent data. Therefore, a feature association algorithm is proposed, which analyzes the correlation between data features and performs feature selection on the initial dataset to prevent overfitting of feature extraction models.

(ii) In this paper, an improved feature extraction model–the deep-related sparse autoencoder (DRSAE) model—is proposed combined with the feature association algorithm. The model uses multiple hidden layers to extract features, and the sparsity of weights and neuron activation degrees in hidden layers is limited by L_1 regularization and Kullback–Leibler divergence (KLD) to increase the accuracy of encoder feature extraction.

The remainder of this article is divided into the following four sections. The second section analyzes the existing research on autoencoder-based feature extraction methods. The third section describes the proposed feature association algorithm and the DRSAE model, and elaborates on the principle and structure of the model. In the fourth section, ATM dataset is used for feature association analysis and feature extraction experiments, and the support vector machine (SVM) classifier is combined to evaluate the classification effect of the feature extraction network model. Finally, the superiority of the proposed DRSAE model in the field of ATM feature extraction is proved by the simulation experiment of ATM system. In addition, the fifth section summarizes the contributions and deficiencies of this paper, and puts forward the next work.

2. Related Work

ATM system is vulnerable to network attacks due to various types of business and multi-source heterogeneity of equipment operation data. And the overall equipment architecture is complex. Once the key equipment has problems, it will affect the stability of the whole ATM system. Therefore, the information security of ATM system has received extensive attention in aviation field. The Federal Aviation Administration (FAA) and other stakeholders are experimenting with data interconnection sharing. ATM requires the highest level of resilience, multi-level redundancy, and dynamic environment. This fragile environment brings a new attack medium to ATM. Moreover, the connection of some legacy systems of ATM system to the network will produce serious threats and vulnerabilities [4]. Aiming at the information security research of ATM system, Chivers[5] proposed a control consistency as a method to check and exchange control set information in the management system and studied the systematic defects of ATM security. The ADS-B monitoring system in ATM system is vulnerable to security threats. Leonardi et al. [6] proposed to use the carrier phase of the transmitter as features to determine the type of aircraft, thereby distinguishing legitimate and false information. At present, the international research on feature extraction of CNS equipment security situation in ATM system is still in the initial research stage, and a systematic theoretical system has not been formed. Therefore, this paper is devoted to study the feature extraction method of ATM system security situation.

The traditional linear feature extraction method is to map the initial feature to a lower dimension through linear projection. This method can no longer meet the requirements of processing multi-dimensional data generated by complex information exchanges in the era of big data. After dimensionality reduction, it can only ensure that the mathematical relationship between data remains relatively unchanged, but the nonlinear dimensionality reduction method can keep the feature information of complex nonlinear data while dimensionality reduction, ensuring that the essential features of the original data do not change. It belongs to the nonlinear manifold learning method [7]. The nonlinear dimensionality reduction method based on deep learning does not use manual and expert knowledge for feature extraction, but uses hyperparameter adaptive feedback training to obtain the optimal model. Autoencoders show significant advantages in deep learning-based feature extraction, and many scholars have improved them to improve the accuracy of feature extraction models. Xi et al. [8] used a feature correlation-based autoencoder (AE) anomaly detection method to reduce the impact of correlations between data and constructed a data association model combined with graph neural network to fuse sample features and increase the detection precision of model. Liu et al. [9] proposed a batch normalized stacked sparse autoencoder (SSAE) method to diagnose equipment faults, which has better detection ability than other methods. Lee et al. [10] proposed a feature extraction method based on deep unsupervised sparse autoencoder (SAE) for data classification, which improved the classification performance and detection speed, but the performance of sparse classes was worse than that of other classes. Xu et al. [11] proposed a deep belief sparse autoencoder (DBSAE), which captured features of label-free dissolved gas analysis (DGA) raw data, and a supervised trained back propagation

network is used to implement transformer fault diagnosis. Marir et al. [12] proposed a new stack denoising SAE method, which was implemented by using spark-based iterative simplification paradigm to improve detection performance and algorithm efficiency. Tang et al. [13] considered the full structure of features in the AE, constraining the AE by adding low-rank properties and Laplace operator structure on features. For solving the problems of lack of fault dataset, fuzzy features, interaction between components, and coupling of fault features, Xie et al. [14] proposed an improved SAE method combined with a multi-level denoising strategy to diagnose electromagnetic interference faults. The method uses a relational constraint to limit the relationship between electromagnetic interference data, and multi-level denoising is carried out for fault data to enhance the ability of feature expression. Feng and Duarte [15] proposed an unsupervised feature selection method based on graph and AE to perform the spectrogram analysis and the feature extraction. Han et al. [16] proposed an ensemble autoencoder (EAE) model based on SAE and denoising autoencoder (DAE), and used convolutional neural network (CNN) pooling layer to control model overfitting, but this model only uses a single hidden layer, and the feature extraction accuracy is not high. Miao et al. [17] designed sparse representation convolutional autoencoder (SRCAE) model for making a fault analysis of the equipment, which made the performance of SRCAE better than of deep neural network (DNN), but the proposed model did not consider the multi-source heterogeneity of the equipment data.

The operation data of ATM system equipment have the characteristics of multi-source, complex, and feature correlation. In order to transform the complex and redundant high-dimensional data into low-dimensional data which is easy to study, and realize the situation awareness of the ATM system to provide decision basis for managers, it is indispensable to study the situation feature extraction method suitable for ATM system.

3. Materials and Methods

The feature selection method can reduce the data dimension, and the output data dimension is only a subset of the input data dimension. However, the AE is different from this method, and by learning the potential relationship between data, the AE reconstructs the data with similar structure to the input data, which has the characteristics of input and output data correlation, output data loss, and unsupervised and automatic learning. The AE uses artificial neural network to reduce the dimension of data by minimizing the reconstruction loss, which does not involve dataset annotation, thus reducing the workload of manual or automatic dataset annotation.

3.1. Autoencoder. A typical AE mainly consists of encoder, hidden unit, and decoder. The encoder and decoder are connected by hidden unit of neural network. The AE uses back propagation method to set the target output to a

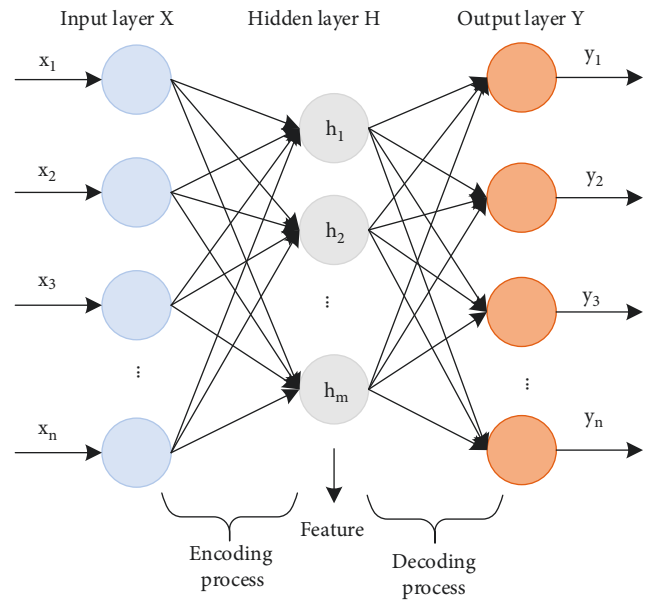

Figure 1: The structure of AE.

structure approximately equal to the input. Feature extraction of data mainly focuses on the structure of the encoder, and data can be converted to a new feature representation after encoder network training, namely, the feature obtained by dimensionality reduction [18]. In addition, the output results similar to the input data are obtained by reconstructing the features through the decoder. Figure 1 shows the architecture of an AE.

The encoder in an AE is the mapping from the input layer $\mathbf{X} = \{x_1, x_2, x_3, \ldots, x_n\}$ to the hidden layer $H = \{h_1, h_2, \ldots, h_m\}$, that is, the encoding process. Among them, m is the quantity of neurons in $\mathbf{H}$, and n is the feature dimension of X. The function $f(x)$ of $\mathbf{H}$ can be formulated as follows:

$$h = f(x) = a_e(w_e x + b_e), \tag{1}$$

where the function $a_e(^*)$ is the nonlinear activation function of the encoder, which is used to convert the input data into the output signal through nonlinear transformation, and w_e and b_e, respectively, represent the weight and deviation value of the $\mathbf{X}$ mapped to the $\mathbf{H}$.

The decoder in the AE is a reconstruction mapping from the hidden layer $H = \{h_1, h_2, \ldots, h_m\}$ to the output layer $Y = \{y_1, y_2, y_3, \ldots, y_n\}$, that is, the decoding process. The function $g(h)$ of $\mathbf{Y}$ is shown as follows:

$$y = g(h) = a_d(w_d h + b_d), \tag{2}$$

where $a_d(^*)$ is the nonlinear activation function of the decoder, and w_d and b_d, respectively, represent the weight and deviation value of the $\mathbf{H}$ mapped to the $\mathbf{Y}$.

The above values of w_e, b_e, w_d, and b_d can be obtained by training the whole AE model. The AE requires training to make the reconstruction error between $\mathbf{X}$ and $\mathbf{Y}$ close to zero, so as to obtain the feature representation of the hidden layer $\mathbf{H}$ that best represents the relationship between the input $\mathbf{X}$ data, so that $X \approx Y$ as much as possible. Mean

squared error (MSE) [19] can represent the similarity between **X** and **Y**. MSE is shown as follows:

$$L_{mse} = \frac{1}{n}\sum_{i=1}^{n}\left\|x_i - y_i\right\|^2, \tag{3}$$

where x_i and y_i are data items of **X** and target output **Y**, respectively, and n is the quantity of data units of **X** and **Y**.

3.2. Feature Association Algorithm. The data of ATM operation are multi-source and heterogeneous, the complexity of data space is high, the dependence between data features is strong, and the correlation between data attributes is prominent. During network training, due to the interaction between features, some data attributes may play a greater role in the overall relationship of data than others, which will affect the features extracted by the encoder to a certain extent, making it impossible to fully represent the potential rules between data. This paper uses the method based on data attributes to establish the feature association model, so as to achieve the purpose of decoupling the attributes and features, and make the attribute association clearly expressed [20].

As some attribute features interact with each other, and in order to accurately represent data attributes, each attribute is regarded as an independent variable in this paper, and the established feature association model based on data attributes is shown in Figure 2.

First, let the original feature set of input data X be $T = \{t_1, t_2, t_3, ..., t_k\}$, and k is the total quantity of attributes of **X**, that is, the initial data feature dimension. Secondly, the feature set T is divided according to the correlation of the actual meaning of attributes, which is further expressed as $T = \{R_1, R_2, ..., R_n\}$, n is the quantity of attribute types that divide the data feature set **T**, $\mathbf{R}_i\,(i = 1, 2, ..., n)$ represents a class of features related to attributes, and then $\mathbf{R}_1 = \{t_1, t_2, ..., t_{p_1}\}$, $\mathbf{R}_2 = \{t_{p_1+1}, t_{p_1+2}, ..., t_{p_1+p_2}\}$, ..., $\mathbf{R}_n = \{t_{p_1+p_2+...+p_{n-1}+1}, t_{p_1+p_2+...+p_{n-1}+2}, ..., t_{p_1+p_2+...+p_{n-1}+p_n}\}$. Among them, $p_1 + p_2 + ... + p_{n-1} + p_n = k$, $p_1, p_2, ..., p_n$ are the number of attributes of $R_1, R_2, ..., R_n$ types of features, respectively.

Taking the features t_1 and t_2 in the $\mathbf{R}_1$ class as an example, the degree of correlation between t_1 and t_2 is calculated by calculating the Euclidean distance between t_1 and t_2. The correlation [21] between features t_1 and t_2 can be formulated as follows:

$$c_{12} = \begin{cases} \dfrac{\sum_{i=1}^{n'}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n'}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n'}(y_i - \overline{y})^2}}, & \exists x_i \neq 0 \wedge y_i \neq 0, \\ 0, & \forall x_i = 0 \vee y_i = 0, \end{cases} \tag{4}$$

where x_i is the data item in the feature t_1 variable, y_i is the data item in the feature t_2 variable, and n' is the total number of input data items. The correlation coefficient matrix $\mathbf{C}_1$ of the features in $\mathbf{R}_1$ is shown as follows:

$$C_1 = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p_1} \\ c_{21} & c_{22} & \cdots & c_{2p_1} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p_1 1} & c_{p_1 2} & \cdots & c_{p_1 p_1} \end{bmatrix}, \tag{5}$$

where p_1 is the number of attributes of the R_1 class features. The attributes of the data are completely related to themselves, so in (5) above, $c_{11}, c_{22}, ..., c_{p_1 p_1} = 1$, and the correlation coefficient matrix $\mathbf{C}_1$ of $\mathbf{R}_1$ can be rewritten as follows:

$$C_1 = \begin{bmatrix} 1 & c_{12} & \cdots & c_{1p_1} \\ c_{21} & 1 & \cdots & c_{2p_1} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p_1 1} & c_{p_1 2} & \cdots & 1 \end{bmatrix}. \tag{6}$$

The range of results obtained by calculating the correlation c is [−1, 1]. As the value of c approaches 1, it means that the two characteristic variables are positively correlated, when the c tends to −1, it means that the two characteristic variables are negative correlated, and when the c tends to 0, it means that the two feature variables do not influence each other.

The correlation coefficient matrix **C** of each type of feature is obtained by performing correlation analysis on n types of features, and a threshold value is set. Except the diagonal correlation elements of the **C**, when the absolute values of other correlation elements in the **C** are greater than or equal to this threshold, the two features of the correlation element are considered to have a strong correlation. Comparing the mean and variance of the two feature data items, since the size of the variance determines the degree of influence of this feature on the overall data, the feature with small variance is selected to be deleted; that is, delete the features that have less impact on the overall data to reduce the feature dimension.

After the correlation analysis is performed on each type of feature, the correlation analysis of the whole data needs to be performed again. If the obtained correlation coefficient matrix shows that there is correlation between attribute features in different categories, take feature t_1 in R_1 class and feature t_{p1+1} in $\mathbf{R}_2$ class as examples, that is, $\exists t_1 \in \mathbf{R}_1, t_{p_1+1} \in \mathbf{R}_2 \Rightarrow c_{1(p_1+1)} \neq 0$. When $|c_{1(p_1+1)}|$ is greater than or equal to the threshold, compare the attribute numbers p_1 and p_2 of $\mathbf{R}_1$ class and $\mathbf{R}_2$ class; when $p_1 < p_2$, it indicates that the feature t_{p1+1} of $\mathbf{R}_2$ class has more attributes than the feature t_1 of $\mathbf{R}_1$ class. In order to maintain a relative balance in the number of attributes of various features, the feature t_{p1+1} of $\mathbf{R}_2$ class is deleted. When $p_1 > p_2$, the principle is the same. When $p_1 = p_2$, deleting the feature t_1 of $\mathbf{R}_1$ class or the feature t_{p1+1} of $\mathbf{R}_2$ class has the same effect.

The feature association algorithm processes the data before feature extraction, which is equivalent to feature selection on the data, and alleviates the influence of data correlation that affects the accuracy and robustness of feature learning.

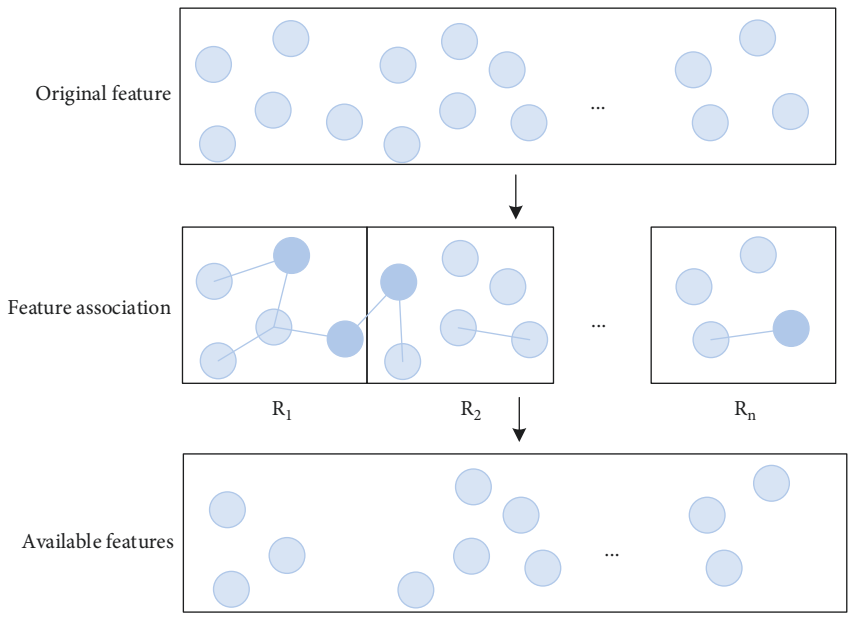

Figure 2: Feature association model.

3.3. DRSAE Feature Extraction Model. In the training process of traditional AE, MSE may be too small, which leads to overfitting of the model, weak generalization ability of the network, and inability to learn important data features effectively. When the dimension m of the hidden layer is expanded to a range greater than or the same as the dimension n of the input layer, it may happen that the output of the AE is equal to the input, and the SAE adds a sparsity penalty constraint to the loss function of the encoding network, so that the encoder can obtain high-dimensional and deeper feature representation, and to some extent, the generalization ability of the encoder is improved by limiting the weight $\mathbf{W}$ of the hidden layer to the input layer [22].

SAE generally uses L_1 regularization and sigmoid function as encoder constraint and hidden layer activation function, respectively. According to the special properties of the sigmoid activation function, the original optimal solution will have different offsets. In the end, some elements of the hidden layer output are close to 0; that is, the elements are in the inactive state, and some elements are close to 1; that is, the elements are in the active state, which makes the encoding network sparse and avoids overfitting of the model. L_1 regularization constrains the weight vector $\mathbf{w}$, and the calculation equation is as follows:

$$L_1 = \lambda \sum_{i=1}^{n} |w_i|, \tag{7}$$

where λ is the sparsity constraint, which is used to control the degree of regularization, and w_i is the weight of the hidden layer $\mathbf{H}$ to each sample of the input $\mathbf{X} = \{x_1, x_2, x_3, ..., x_n\}$. After L_1 regularization restriction, the loss function of the encoder is shown as follows:

$$L_{1-AE} = L_{\text{mse}} + L_1 = \frac{1}{n} \sum_{i=1}^{n} \|x_i - y_i\|^2 + \lambda \sum_{i=1}^{n} |w_i|. \tag{8}$$

In addition to using the L_1 regularization to constrain the loss function, the relative entropy, namely, the Kullback–Leibler divergence (KLD), can also be used as a penalty term to asymmetrically measure the difference between $\mathbf{X}$ and the target $\mathbf{Y}$ probability distribution to limit the sparsity of the network, and KLD [23] is defined as follows:

$$D_{kl}(p_x \| p_y) = \sum_{p_x} p_x(x_i) \cdot \log \frac{p_x(x_i)}{p_y(y_i)}, \tag{9}$$

where p_x and p_y are Bernoulli distributions of random variables x and y, and $p_x(x_i)$ and $p_y(y_i)$ are the probability distribution functions of x_i and y_i. In SAE, the average activation degree a_j of the jth neuron in $\mathbf{H}$ to $\mathbf{X} = \{x_1, x_2, x_3, ..., x_n\}$ is shown as follows:

$$a_j = \frac{1}{n} \sum_{i=1}^{n} \left| a_{h(j)}(x_i) \right|, j \in \{1, 2, 3, ..., m\}, \tag{10}$$

where $h(j)$ is the jth component of the matrix vector of $\mathbf{H}$ and $a_{h(j)}$ is the overall activation degree of the jth neuron in $\mathbf{H}$; when the input sample is $X = \{x_1, x_2, x_3, ..., x_n\}$, $\mathrm{a_{h(j)}(x_i)}$ indicates the activation degree of the jth neuron in $\mathbf{H}$; when the input data are x_i, its value depends on the weight and deviation in (1).

To ensure the sparsity of the network, neurons in $\mathbf{H}$ need to be in an inhibited state most of the time; that is, the quantity of neurons in the active state is much smaller than the quantity of neurons in the inactive state. Let a be the expected average activation degree of neurons in $\mathbf{H}$ for the input sample, which is a value close to 0 (usually set to 0.05). In an ideal state, the actual average activation degree a_j of each neuron should be equal to the sparsity parameter a, and the KLD of a_j and a is shown as follows:

$$D_{kl}\left(a\|a_j\right) = a \cdot \log\frac{a}{a_j} + (1-a) \cdot \log\frac{1-a}{1-a_j}, \quad j \in \{1, 2, 3, ..., m\}. \tag{11}$$

The smaller $D_{kl}(a\|a_j)$ is, the smaller the difference between input distribution and output distribution of SAE is, and the smaller the loss of encoder network is. The target loss function of SAE is shown as follows:

$$L_{\text{KLD}-AE} = L_{\text{mse}} + \mu \sum_{j=1}^{m} D_{kl}\left(a\|a_j\right) = \frac{1}{n}\sum_{i=1}^{n}\left\|x_i - y_i\right\|^2 + \mu \sum_{j=1}^{m} D_{kl}\left(a \parallel a_j\right), \tag{12}$$

where μ is the penalty constraint parameter imposed on a_j when a_j is not equal to λ, that is, the penalty factor of the penalty term $D_{kl}(a\|a_j)$, which represents the sparse degree of neurons in $\mathbf{H}$.

When the encoder uses the sigmoid activation function, the range of the output feature values of $\mathbf{H}$ is between (0, 1). The loss function can be calculated using KLD, but it is prone to gradient saturation. When the encoder uses the rectified linear unit (ReLU) activation function, since there is no gradient vanishing problem, the range of output feature values of the hidden layer is $[0, +\infty)$, and the denominator may be 0 when calculating KLD. Therefore, the KLD cannot be used to constrain the model, and only L_1 regularization can be used to make the encoder learn the sparsity characteristics of the data.

In addition, when training the encoder network, the setting of hyperparameters, namely, sparse parameters, will affect the sparsity of the network target output. The stochastic gradient descent (SGD) algorithm and adaptive moment estimation (Adam) optimizers are usually used to train sparse parameters iteratively through back propagation to determine the optimal parameter values, thus minimizing the loss function.

In this paper, SAE is used for extracting features from the multi-dimensional data of ATM system and converting the equipment operation data from multi-dimensional space to one-dimensional space or a dimension that is favorable for subsequent research. By improving the hidden layer of the original AE, SAE regards some nodes of each unit in the hidden layer as inactive states and only studies the correlation and characteristics of other nodes in this layer [24], which can improve the accuracy of feature extraction of the encoder.

Increasing the number of hidden layers in SAE can make the encoder learn more useful hidden structures and representations of data, and make SAE become a deep sparse autoencoder (DSAE), which is more accurate than models using SAE with a single hidden layer in terms of system feature extraction. Therefore, based on the coupling problem between features of ATM system, this paper increases the number of hidden layers in SAE for improving the data quality of input network and training the AE layer by layer. Finally, a DSAE feature extraction model is formed. The structure of DSAE is appeared in Figure 3.

DSAE uses MSE to measure the similarity between the input and output of the autoencoder, and uses L_1 regularization to impose regularization constraints on the encoder, which makes the encoder generate a sparse weight matrix. Then, combined with KLD, the activation degree of neurons in H of the encoder is limited to increase the accuracy of feature extraction model. Since the encoder calculates the gradient value by averaging the partial derivatives, which only need to sum the partial derivatives of all terms, the loss function can also be directly expressed by adding together. According to (8) and (11), the objective loss function of DSAE is shown as follows:

$$L_{\text{DSAE}} = L_{\text{mse}} + L_1 + \mu \sum_{j=1}^{m} D_{kl}\left(a\|a_j\right) = \frac{1}{n}\sum_{i=1}^{n}\left\|x_i - y_i\right\|^2 + \lambda \sum_{i=1}^{n}\left|w_i\right| + \mu \sum_{j=1}^{m} D_{kl}\left(a\|a_j\right). \tag{13}$$

In addition, the data in this paper are sparse, the features with low frequency need to be updated more frequently, and the learning rate will gradually decrease with the number of updates. Therefore, the Adam optimizer is used as the optimizer of the encoder parameters to calculate adaptive learning rate for each parameter, and the optimal parameter combination is found through back propagation during training.

On the basis of DSAE, the proposed feature association algorithm is integrated into DSAE to form DRSAE feature extraction model. The established DRSAE model is shown in Figure 4.

Since the DRSAE model integrates the feature association algorithm, the initial input of the model has changed, and the loss function of the improved DRSAE model has also changed. According to (13), the loss function of DRSAE is as follows:

$$L_{\text{DRSAE}} = L_{\text{mse}}{}' + L_1' + \mu \sum_{j=1}^{m'} D_{kl}\left(a\|a_j\right) = \frac{1}{s}\sum_{i=1}^{s}\left\|x_i - y_i\right\|^2 + \lambda \sum_{i=1}^{s}\left|w_i\right| + \mu \sum_{j=1}^{m'} D_{kl}\left(a\|a_j\right), \tag{14}$$

where $x_i \in \mathbf{X}'$, $y_i \in \mathbf{Y}'$, $\mathbf{X}'$ and $\mathbf{Y}'$ are the input dataset and output dataset of the model after feature association selection, respectively, $X' = \{x_1, x_2, x_3, ..., x_s\}$, $\mathbf{X}' \subseteq \mathbf{X}$, $Y' = \{y_1, y_2, y_3, ..., y_s\}$, $\mathbf{Y}' \subseteq \mathbf{Y}$, s is the dimension of the input data and output data of the model after feature association selection, m' is the number of neurons in the hidden layer of the model after feature association selection, $s \le n$.

The algorithm steps of DRSAE feature extraction model are as follows:

(i) Normalized dataset.

(ii) Perform feature correlation analysis on the data, obtain the correlation coefficient matrix of each type of feature by calculating correlation, use the threshold to compare the correlation, and delete

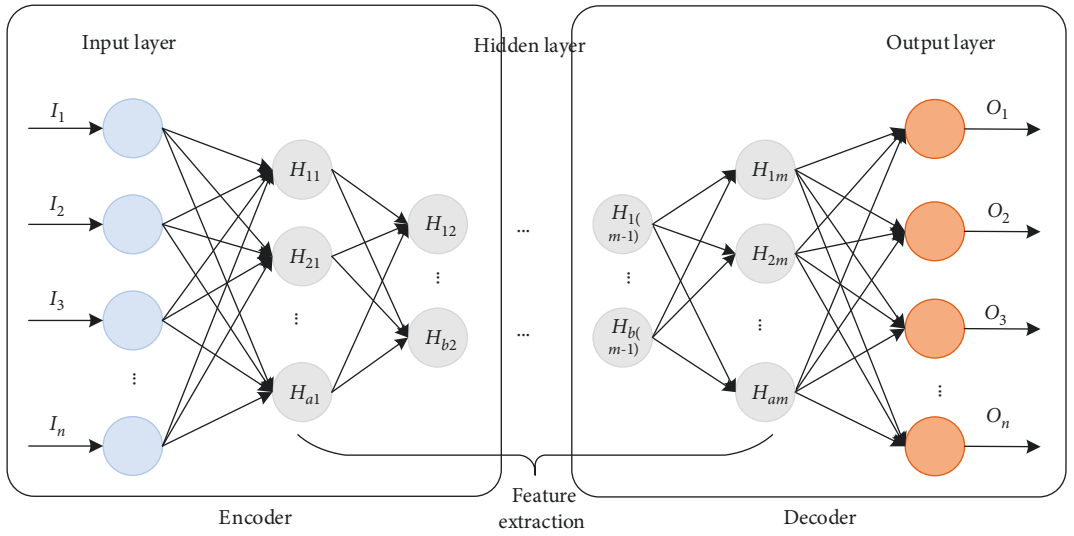

Figure 3: The structure of DSAE.

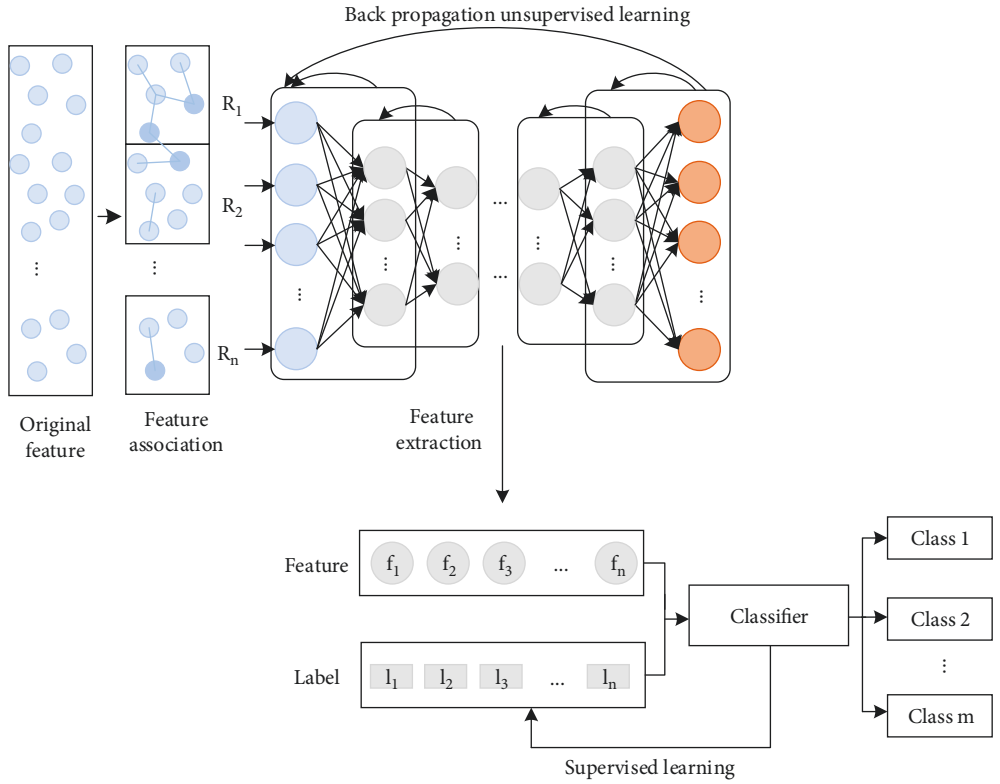

Figure 4: DRSAE feature extraction model.

the features that have little impact on the overall data to obtain a new dataset.

(iii) Divide the training set and the testing set on the basis of the size of the dataset, and input the training set into DSAE.

(iv) Train the encoder models one by one, and unsupervisedly learn the features of the data through the fully connected layer of the autoencoder to make the contrast loss between the reconstructed output and the input tend to zero as much as possible.

(v) Use the encoder output weight of the last trained associative SAE as the encoder input weight of the current associative SAE, and train the current associative SAE.

(vi) Connect the trained encoder layers of each SAE to form a DRSAE, initialize parameters of the entire DRSAE model using the previously trained model

parameters, and carry out global optimization by Adam optimizer.

(vii) After training the model, extract the output weight of the last hidden layer of the encoder, that is, the extracted final features, input these features and their corresponding labels into a set classifier, and carry out supervised learning through labeled sample data.

(viii) After training the classifier, get the classification result of the model.

4. Results and Discussion

In this paper, Windows 7 (64 bit) operating system, Intel Core processor is used as the experimental environment. Experiments are carried out on the equipment operation data of the ATM system through the Python language. The Keras framework in deep learning is used to train the feature extraction and classification model.

4.1. Parameter Description. The DRSAE method in this paper first performs feature association analysis on the data, uses the unsupervised learning method to learn the data features through the sparse autoencoder, adds an SVM classifier to the last layer of the encoder, and performs supervised learning combined with the data labels. The whole model actually is a semi-supervised learning model. Among them, the activation function of the hidden layers of the model adopts the sigmoid nonlinear activation function to perform nonlinear transformation, and the SVM uses the radial basis function (RBF) as the kernel function to classify the normal and abnormal data. The values of the model parameters used in the experiments are set as shown in Table 1.

The threshold in the feature association algorithm is set based on empirical judgment on the strength of the ATM data feature association, which has certain subjective factors. Usually, the threshold is set to 0.9 in experiments. The value of a is close to 0. In an ideal state, the actual average activation degree a_j of each neuron should be equal to the sparsity parameter a. In experiments, it is generally set to 0.05. The settings of λ, μ, and the dimension of the hidden layer are explained in the experiments below. Since the setting of SVM classifier has no impact on our study and is only used to evaluate the performance of feature extraction model, the penalty coefficient of SVM objective function is set as a fixed 0.5.

4.2. Data Information and Preprocessing. According to the regulations for operation and maintenance of China Civil Aviation Communication, Navigation and Surveillance System [25], this paper conducts simulation experiments on the operation data of ATM system equipment. Select three representative pieces of data from the ATM system equipment operation dataset, as shown in Table 2, including planned total working hours (PH), normal working hours (NH), normal operation rate (NOR), total number of equipment (TN), number of faulty equipment (NF), equipment intact rate (EIR), number of accidents (NA), number of serious errors (NS), and number of general errors (NG). A piece of data without accidents and errors is regarded as normal data and is recorded as 0, and a piece of data with accidents and errors is regarded as abnormal data and is recorded as 1.

Table 1: Parameter settings.

Parameter	Value
Feature association threshold	0.9
a	0.05
λ	0.0005
μ	1
Dimensional change of encoding layer 1	$8 \longrightarrow 7$
Dimensional change of encoding layer 2	$7 \longrightarrow 4$
Dimensional change of decoding layer 1	$4 \longrightarrow 7$
Dimensional change of decoding layer 2	$7 \longrightarrow 8$
Penalty coefficient of SVM objective function	0.5

The calculation methods of NOR and EIR are shown as follows:

$$\mathrm{NOR} = \frac{\mathrm{NH}}{\mathrm{PH}} \times 100\%, \tag{15}$$

$$\mathrm{EIR} = \frac{\mathrm{TN} - \mathrm{NF}}{\mathrm{TN}} \times 100\%. \tag{16}$$

The normalization method used for data preprocessing [26] is shown as follows:

$$x_{\mathrm{scale}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \cdot (\max - \min) + \min, \tag{17}$$

where max is the maximum value of the scope of data normalization, min is the minimum value, and the scope of this experiment is set to [0, 1].

4.3. Feature Association Analysis. The features of the ATM operation dataset include 9 attributes, which are separated into three classes, namely, running time feature (including 3 attributes), running quantity feature (including 3 attributes), and running error features (including 3 attributes), which are recorded as $\mathbf{R}_1$, $\mathbf{R}_2$, and $\mathbf{R}_3$ features, respectively. Feature association analysis was carried out for these three types of features, respectively, and the heat maps of correlation coefficient matrix of each feature class are shown in Figure 5.

According to the feature association algorithm, the threshold is set to 0.9, when the absolute value of the correlation between two feature attributes is greater than or equal to 0.9, the feature is selected, and the data item is deleted. It can be seen from Figure 5 that the relationship between the $\mathbf{R}_1$ class and $\mathbf{R}_3$ class feature attributes is weak, and the absolute values of the correlation are less than the threshold of 0.9. While attribute 2 and attribute 3 in $\mathbf{R}_2$ features have a strong correlation, the absolute value of the correlation is 0.98, higher than the threshold value of 0.9. According to the feature association algorithm, the variance of the two is compared, and the data item of attribute 2 is determined to be deleted. After the above analysis, the number of features of the $\mathbf{R}_1$ category remains unchanged at

Table 2: Equipment operation data information of ATM system.

Data use cases	PH (h)	NH (h)	NOR (%)	TN	NF	EIR (%)	NA	NS	NG	Label
1	2160	2153	99.7	5	0	100	0	0	1	1
2	2155	2100	97.4	6	1	83.3	0	1	1	1
3	2100	2100	100	3	0	100	0	0	0	0

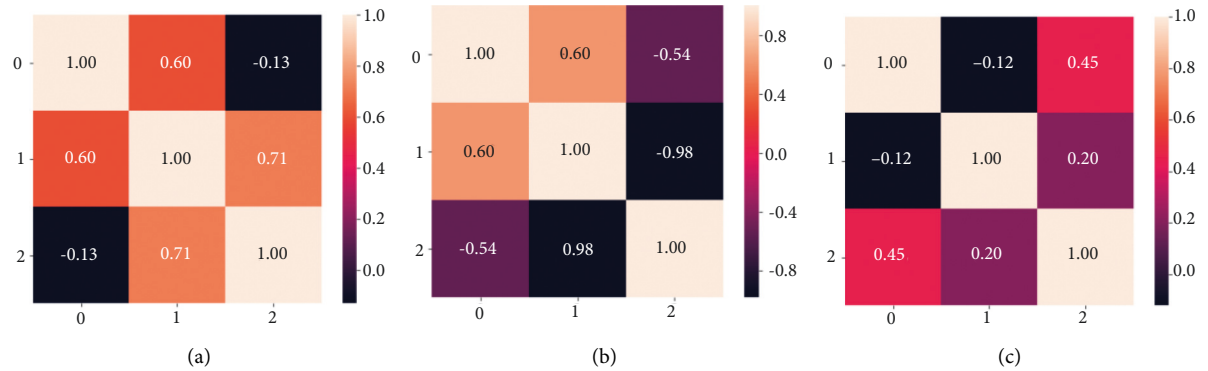

Figure 5: The heat maps of correlation coefficient matrix of each feature class for ATM dataset. (a) The heat map of correlation coefficient matrix of R_1. (b) The heat map of correlation coefficient matrix of R_2. (c) The heat map of correlation coefficient matrix of R_3.

3, the number of features of the $\mathbf{R}_2$ category becomes 2, the number of features of the $\mathbf{R}_3$ category remains unchanged at 3, and the final data dimension is 8.

After performing the above correlation analysis on the features of the $\mathbf{R}_1$, $\mathbf{R}_2$, and $\mathbf{R}_3$ categories and eliminating one of the feature attributes with strong correlation, the overall correlation of the obtained new data is analyzed. The correlation coefficient matrix of the overall data is shown in Figure 6.

In Figure 6, the absolute values of the correlation between attributes of $\mathbf{R}_1$, $\mathbf{R}_2$, and $\mathbf{R}_3$ features are all less than 0.9, so the feature numbers of $\mathbf{R}_1$, $\mathbf{R}_2$, and $\mathbf{R}_3$ remain unchanged, and the data dimension finally obtained is 8.

4.4. Experimental Results and Analysis. The feature extraction model is used to train and test the ATM dataset, and SVM is combined to judge the data category. All models are set to reduce the input data from 8 dimensions to 4 dimensions. Assume that in the initial experiments, the first encoding layer of DRSAE model reduces the input 8-dimensional data to 6-dimensional space, and the second coding layer reduces the input 6-dimensional data to 4 feature spaces. The dimensional change of the two decoding layers is opposite to that of the encoding layer. Finally, the reconstructed final features are obtained through iterative training.

In the DRSAE model proposed in this paper, in order to select appropriate parameter values, this paper conducted a comparative experiment on how the changes of each parameter affect the experimental results. When the batch size is 32, 64, 128, and 256, respectively, the changes in the classification score of the DRSAE model after feature extraction are shown in Figure 7. When the epochs of the entire DRSAE are 50, 100, 150, and 200, respectively, the

Figure 6: The heat map of overall correlation coefficient matrix of ATM dataset after feature association analysis.

changes in the classification accuracy of the DRSAE model after feature extraction are shown in Figure 8.

It can be seen from Figures 7 and 8 that the batch size and epochs have a certain impact on the classification results. When the batch size is 64 and the epoch is 100, the classification score and accuracy results are the best, 63.89% and 71.36%, respectively.

In DRSAE model, the value of penalty constraint μ is very important for the actual average activation degree a_j of each neuron. Five evaluation indexes are used to evaluate the influence of μ on classification results, including accuracy, precision, true-positive rate (TPR), false-positive rate (FPR), and F-score [27]. The evaluation results are shown in Figure 9.

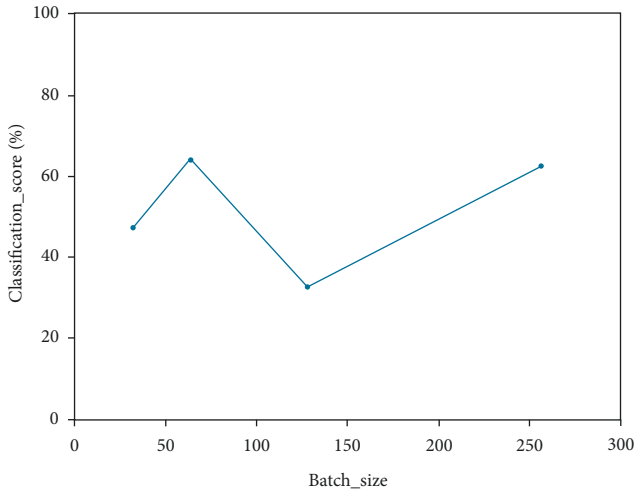

Figure 7: The effect of batch size on classification score.

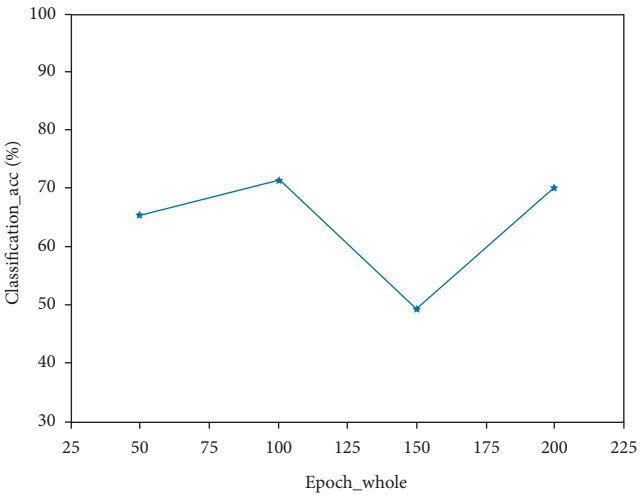

Figure 8: The effect of epochs on classification accuracy.

For the classification evaluation of different feature extraction models, the higher the values of accuracy, TPR, precision, and F-score are, the lower the FPR is, which indicates that the better the classification performance and the higher the accuracy of the model. It can be seen from Figure 9 that when μ is 2, the FPR is the lowest, which is 25.49%, indicating that the proportion of the number of abnormal data classified by the classifier as normal data to the total number of abnormal data in the actual dataset is the smallest. The F-score is a good combination of TPR and precision. When μ is 1, F-score is the highest, which is 78.84%, and the classification accuracy is also the highest, which is 74.37%. FPR is 48.04%, which is slightly higher than when μ is 2. In conclusion, the model has the best performance when μ is 1.

In DRSAE model, the dimension setting of the hidden layer is a necessary thing to be done in this experiment. As the DRSAE model is designed with multiple hidden layers, and the initial dimension of ATM data used in this experiment is 8, there is no need for too many hidden layers to

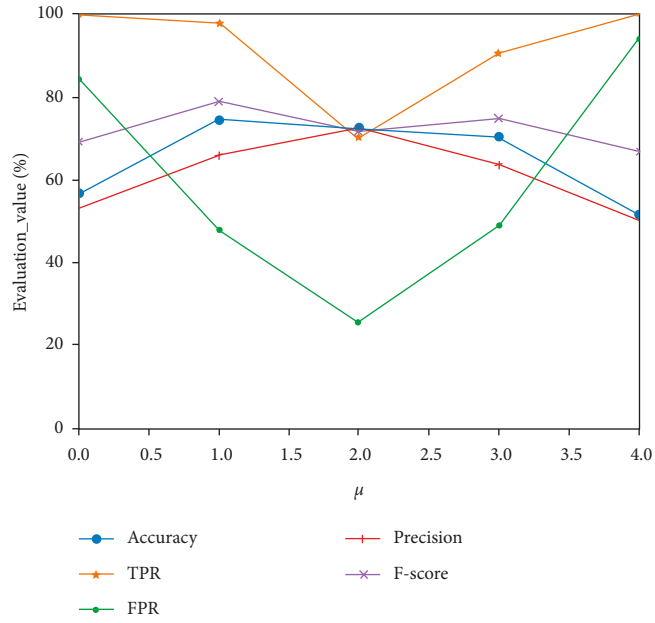

Figure 9: The effect of μ on the classification results.

increase the complexity of feature extraction. Deeper layers may bring overfitting. Therefore, five hidden layers are used for feature extraction, among which the first three layers are the hidden layers of the encoder and the last three layers are the hidden layers of the decoder. ROC and AUC evaluation indexes in dichotomies were used to determine the results. ROC curves corresponding to different dimensions of hidden layers are shown in Figure 10.

Due to the area between ROC curve and x coordinate axis, namely, AUC, the larger the AUC is, the better the model classification performance is. As can be seen from Figure 10, when the dimensions of the three hidden layers of the encoder are 8, 7, and 4, respectively, the AUC is the largest, which is 0.89. Therefore, the dimension of the hidden layers of the encoder is determined to be (8, 7, 4).

In the DRSAE model, the influence of λ controlling the degree of regularization on the experimental results is shown in Figure 11.

As can be seen from Figure 11, when $\lambda = 0.0005$, the classification results have the best performance.

In addition, the DRSAE model can be regarded as using three methods, namely, the feature association method, the L_1 regularization method, and the KLD method. The ablation experiment on the model is shown in Table 3.

It can be seen from Table 3 that when the feature association algorithm is not used, the classification accuracy of the model is less than 70%. When the feature association algorithm is added, the classification accuracy of the model is improved because the association analysis is performed on the features first and the feature selection is realized. Compared with only using the L_1 regularization method, only using the KLD method has a greater effect on the model. And after adding the feature association algorithm,

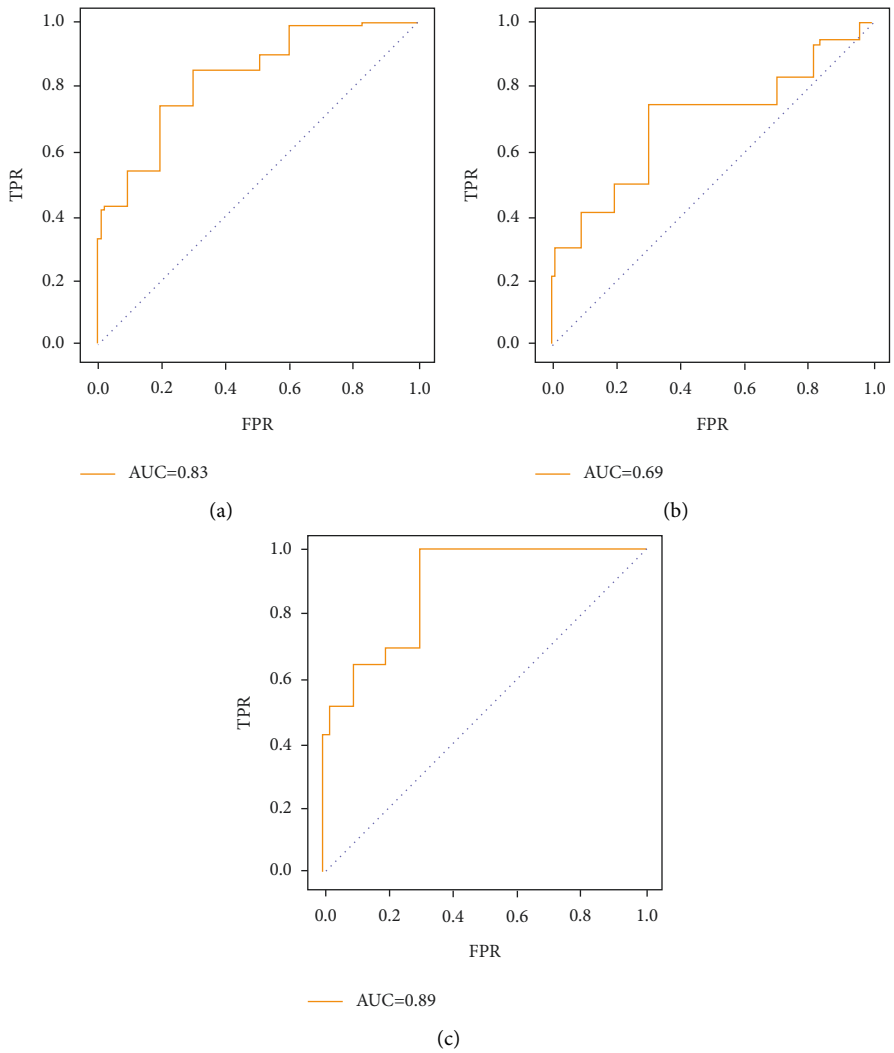

Figure 10: ROC curves corresponding to different dimensions of hidden layers. (a) The dimension of the hidden layer of the encoder = (8, 5, 4) (AUC = 0.83). (b) The dimension of the hidden layer of the encoder = (8, 6, 4) (AUC = 0.69). (c) The dimension of the hidden layer of the encoder = (8, 7, 4) (AUC = 0.89).

the KLD method has a greater effect on the model than the L_1 regularization method. The DRSAE model combines three methods, and the model classification accuracy is greatly improved. Therefore, all three methods are important to the DRSAE model, and none of them are indispensable.

The DRSAE model in this paper was compared with the typical linear dimension reduction methods, such as principal component analysis (PCA) [28] model, and the nonlinear dimension reduction methods, such as AE [29] and SAE [30] models by using parameter settings in Table 1, and SVM classifier was used to classify data. The feature extraction time and classification time of each model are shown in Table 4.

The air traffic control system operation data used in this experiment are simulation data, and the amount of data is not large, so the processing time of all models is very fast. In Table 4, the PCA model can only reduce the dimension linearly, the model is relatively simple, and the feature extraction time is the shortest. The DRSAE model proposed in this paper takes the longest time, because the DRSAE model performs nonlinear dimensionality reduction on the data, and the model is more complex. Moreover, the classification time of DRSAE model is 0.005 s, which is in the middle level. Compared with AE and SAE models, this model increases the number of hidden layers and regularization limits, the operation of feature extraction is more complex, and the classification time is slightly longer, but the time is very short, less than 1 s, which can be ignored.

The DRSAE model is compared with PCA, AE, and SAE models, and the classification evaluation results of all models to the testing set are shown in Figure 12.

As can be seen from Figure 12, for the operation data of ATM equipment, since the PCA feature extraction method of linear dimension reduction can only perform linear transformation, the model classification accuracy is the lowest, which is 63.32%. Nonlinear dimensionality reduction methods such as AE and SAE can perform both linear transformation and nonlinear transformation, and their

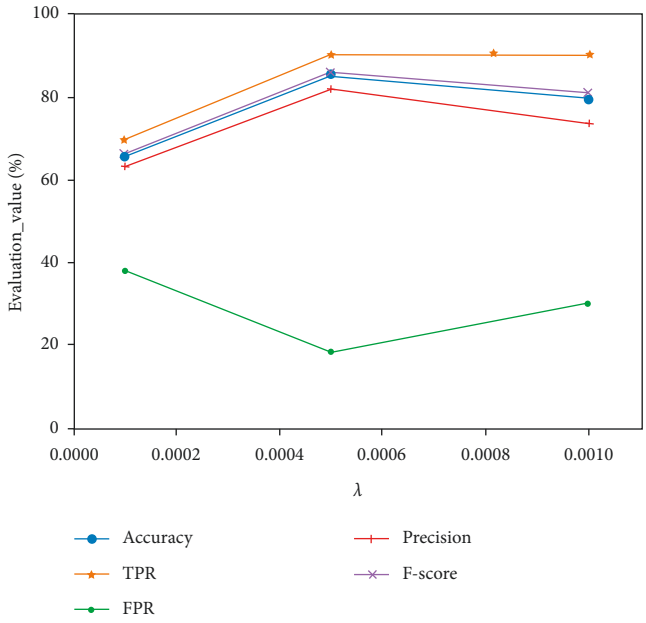

Figure 11: The effect of λ on the classification results.

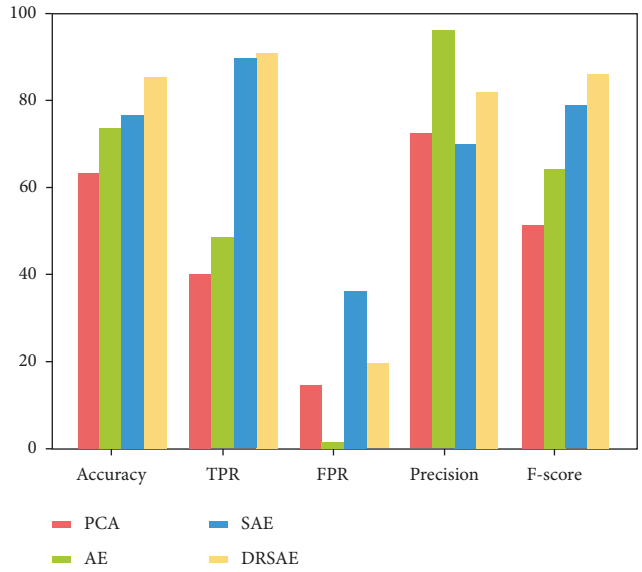

Figure 12: Classification evaluation results of different models on ATM operation testing set.

Table 3: Ablation experiment of DRSAE model.

Method	Classification accuracy (%)
L_1	60.00
KLD	68.89
L_1, KLD	62.78
Feature association, L_1	68.33
Feature association, KLD	70.56
Feature association, L_1, KLD	84.42

Table 4: Comparison of feature extraction time and classification time for each model.

Model	Feature extraction time (s)	Classification time (s)
PCA	0.120	0.010
AE	2.390	0.040
SAE	0.920	0.003
DRSAE	5.800	0.005

model classification accuracy has been improved accordingly, indicating that nonlinear dimension reduction method can extract more effective feature representation for the current dataset with complex feature space. However, the nonlinear dimension reduction model DRSAE in this paper adds sparsity restrictions and the number of hidden layers. The DRSAE model has the highest classification accuracy of 85.43%, which is about 9% higher than SAE model and has the strongest feature extraction performance. For FPR, it can be seen that the classification error rate of DRSAE in this paper is 19.61%, which is in the middle position due to the error of model training. It can be seen that the F-score of the DRSAE model in this paper is the highest, which is 85.85%, indicating that the situation features of the ATM system extracted by the DRSAE method have strong expression ability, and the DRSAE feature extraction model has the peculiarities of high accuracy and strong classification performance, which is convenient for the subsequent situation assessment of ATM system.

5. Conclusions

In this paper, a SAE-based feature extraction method for ATM system is proposed. Because of the correlation problem of ATM system data features, a feature association algorithm is designed, and a deep-related sparse autoencoder (DRSAE) model is established. L_1 regularization and KLD sparsity term are combined to limit the encoder network. After iterative training of a single encoder, the whole encoder is globally optimized. Finally, SVM classifier is combined to evaluate the DRSAE model and other feature extraction models. The experimental results show that, compared with other linear dimension reduction methods, such as PCA and nonlinear dimension reduction methods, such as AE, the DRSAE model considers the correlation between data, increasing the quantity of hidden layers enables the model to extract more expressive and robust features, thus achieving smaller structural loss and lower classification error rates, and the model has stronger performance.

However, this paper does not conduct multi-class evaluation of the model, and the model performance can be further optimized. In the future, more advanced classifiers will be used to achieve multi-classification of situation data, and a new situation assessment model will be proposed to evaluate the situation of ATM system by calculating the situational value, so as to provide more reliable decision-making basis for controllers.

Data Availability

The ATM dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the joint funds of National Natural Science Foundation of China and Civil Aviation Administration of China (U1933108 and U2133203), the National Natural Science Foundation of China (62172418), the Natural Science Foundation of Tianjin China (21JCZDJZ00830), and the Open Fund of Key Laboratory of Airworthiness Certification Technology of Civil Aviation Aircraft (SH2021111907).

References

[1] L. Bogoda, J. Mo, and C. Bil, "A systems engineering approach to appraise cybersecurity risks of CNS/ATM and avionics systems," in *Proceedings of the 2019 Integrated Communications, Navigation and Surveillance Conference*, pp. 1–15, ICNS), Herndon, VA, USA, April 2019.

[2] H. Zhang, Y. Yin, and D. Zhao, "Network security situation awareness model based on threat intelligence," *Journal of Communications*, vol. 42, no. 6, pp. 182–194, 2021.

[3] S. Li, L. Jiang, Q. Zhang, Z. Wang, Z. Tian, and M. Guizani, "A malicious mining code detection method based on multi-features fusion," *IEEE Transactions on Network Science and Engineering*, 2022.

[4] P. Lellek, P. Leydold, I. Vojnoski, and D. Eiar, "Anomaly detection in ATM-grade software defined networks," in *Proceedings of the 2021 Integrated Communications Navigation and Surveillance Conference*, pp. 1–8, ICNS), Dulles, VA, USA, April 2021.

[5] H. Chivers, "Control consistency as a management tool: the identification of systematic security control weaknesses in air traffic management," *International Journal of Critical Computer-Based Systems*, vol. 6, no. 3, p. 229, 2016.

[6] M. Leonardi, L. Di Gregorio, and D. Di Fausto, "Air traffic security: aircraft classification using ADS-B message's phase-pattern," *Aerospace*, vol. 4, no. 4, p. 51, 2017.

[7] Q. Meng, D. Catchpoole, D. Skillicom, and P. Kennedy, "Relational autoencoder for feature extraction," in *Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 364–371, Anchorage, AK, USA, May 2017.

[8] L. Xi, R. Wang, and H. Fan, "Unsupervised deep anomaly detection model based on sample association perception," *Chinese Journal of Computers*, vol. 44, no. 11, pp. 2317–2331, 2021.

[9] X. Liu, Y. Gao, and Y. Yang, "Fault diagnosis based on batch-normalized stacked sparse autoencoder," in *Proceedings of the 2020 39th Chinese Control Conference (CCC)*, pp. 4141–4146, Shenyang, China, July 2020.

[10] J. Lee, J. Pak, and M. Lee, "Network intrusion detection system using feature extraction based on deep sparse autoencoder," in *Proceedings of the 2020 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 1282–1287, Jeju, Korea, October 2020.

[11] Z. Xu, W. Mo, Y. Wang, S. Luo, and T. Liu, "transformer fault diagnosis based on deep brief sparse autoencoder," in *Proceedings of the 2019 Chinese Control Conference (CCC)*, pp. 7432–7435, Guangzhou, China, July 2019.

[12] N. Marir, H. Wang, and G. Feng, "Unsupervised feature learning with distributed stacked denoising sparse autoencoder for abnormal behavior detection using Apache Spark," in *Proceedings of the 2019 IEEE 2nd International Conference on Knowledge Innovation and Invention (ICKII)*, pp. 473–476, Seoul, Korea, July 2019.

[13] K. Tang, K. Xu, Z. Su, W. Jiang, X. Luo, and X. Sun, "Structure-constrained feature extraction by autoencoders for subspace clustering," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 624–632, Seoul, Korea, March 2020.

[14] G. Xie, J. Yang, and Y. Yang, "An improved sparse autoencoder and multilevel denoising strategy for diagnosing early multiple intermittent faults," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 2, pp. 869–880, 2022.

[15] S. Feng and M. F. Duarte, "Graph regularized autoencoder-based unsupervised feature selection," in *Proceedings of the 2018 52nd Asilomar Conference on Signals, Systems, and Computers*, pp. 55–59, Pacific Grove, CA, USA, February 2019.

[16] Y. Han, Y. Ma, J. Wang, and J. Wang, "Research on ensemble model of anomaly detection based on autoencoder," in *Proceedings of the 2020 IEEE 20th International Conference on Software Quality, Reliability and Security (QRS)*, pp. 414–417, Macau, china, December 2020.

[17] M. Miao, Y. Sun, and J. Yu, "Sparse representation convolutional autoencoder for feature learning of vibration signals and its applications in machinery fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 12, pp. 13565–13575, 2022.

[18] B. B. Li, *Anomaly Detection and Classification of Network Traffic Data Based on Self-Coding Neural Network*, Jiangnan university, China, 2021.

[19] J. B. Yao, *Power Load Prediction Based on Deep Sparse Autoencoder Dimension Reduction and Improved Deep Learning GRU Model*, Yanshan university, China, 2021.

[20] H. Fan, F. Zhang, and Z. Li, "Anomalydae: dual autoencoder for anomaly detection on attributed networks," in *Proceedings of the ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5685–5689, Barcelona, Spain, May 2020.

[21] N. Lu, X. Zhang, and N. Ou, "Zero-sample semantic autoencoder based on particle swarm optimization for optimal attribute correlation," *Journal of Electronics and Information Technology*, vol. 43, no. 4, pp. 982–991, 2021.

[22] H. Yang, Z. Zhang, and L. Zhang, "Based on parallel feature extraction and improve BiGRU network security situation assessment," *Journal of Tsinghua University*, vol. 62, no. 5, pp. 842–848, 2022.

[23] H. K. Bhuyan, D. C. Chakraborty, S. K. Pani, and V. Ravi, "Feature and subfeature selection for classification using correlation coefficient and fuzzy model," *IEEE Transactions on Engineering Management*, pp. 2021–2115, Article ID 3065699, 2022.

[24] B. Yan and G. Han, "Effective feature extraction via stacked sparse autoencoder to improve intrusion detection system," *IEEE Access*, vol. 6, pp. 41238–41248, 2018.

[25] Caac, "Regulations for operation and maintenance of China Civil aviation communication, navigation and surveillance system," http://www.caac.gov.cn/XXGK/XXGK/GFXWJ/201511/P020151103347016287809.pdf.

[26] J. Liu, A. Zhang, and Z. Huang, "Optimization dimensionality reduction analysis of CSE-CIC-IDS2018 intrusion detection data set based on machine learning," *Fire Control and Command Control*, vol. 46, no. 7, pp. 155–162, 2021.

[27] S. Li, Y. Li, W. Han, X. Du, M. Guizani, and Z. Tian, "Malicious mining code detection based on ensemble learning in cloud computing environment," *Simulation Modelling Practice and Theory*, vol. 113, Article ID 102391, 2021.

[28] C. Wu, G. Qi, H. Zhao, and Z. Chen, "Feature extraction of cultural gene image based on PCA method," in *Proceedings of the 2020 International Conference on Computer Engineering and Application (ICCEA)*, pp. 860–863, Guangzhou, China, March 2020.

[29] M. George and B. R. Jose, "Crowd panic detection using autoencoder with non-uniform feature extraction," in *Proceedings of the 2018 8th International Symposium on Embedded Computing and System Design ISED*, pp. 11–15, Cochin, India, December 2018.

[30] X. Kong, R. Lin, and H. Zou, "Feature extraction of load curve based on autoencoder network," in *Proceedings of the 2020 IEEE 20th International Conference on Communication Technology (ICCT)*, pp. 1452–1456, Nanning, China, October 2020.