

Research Article

Unsupervised Anomaly Detection Based on Deep Autoencoding and Clustering

Chuanlei Zhang ¹, Jiangtao Liu,¹ Wei Chen ^{2,3}, Jinyuan Shi,¹ Minda Yao ¹,
Xiaoning Yan,⁴ Nenghua Xu,⁴ and Dufeng Chen⁵

¹College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin 300457, China

²School of Mechanical Electronic and Information Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China

³School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China

⁴Softsz Co., Ltd., Shenzhen 518131, China

⁵Beijing Geotechnical and Investigation Engineering Institute, Beijing 100080, China

Correspondence should be addressed to Wei Chen; chenwdavior@163.com

Received 17 June 2021; Revised 8 September 2021; Accepted 20 September 2021; Published 13 October 2021

Academic Editor: Chi-Hua Chen

Copyright © 2021 Chuanlei Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The unsupervised anomaly detection task based on high-dimensional or multidimensional data occupies a very important position in the field of machine learning and industrial applications; especially in the aspect of network security, the anomaly detection of network data is particularly important. The key to anomaly detection is density estimation. Although the methods of dimension reduction and density estimation have made great progress in recent years, most dimension reduction methods are difficult to retain the key information of original data or multidimensional data. Recent studies have shown that the deep autoencoder (DAE) can solve this problem well. In order to improve the performance of unsupervised anomaly detection, we propose an anomaly detection scheme based on a deep autoencoder (DAE) and clustering methods. The deep autoencoder is trained to learn the compressed representation of the input data and then feed it to clustering approach. This scheme makes full use of the advantages of the deep autoencoder (DAE) to generate low-dimensional representation and reconstruction errors for the input high-dimensional or multidimensional data and uses them to reconstruct the input samples. The proposed scheme could eliminate redundant information contained in the data, improve performance of clustering methods in identifying abnormal samples, and reduce the amount of calculation. To verify the effectiveness of the proposed scheme, massive comparison experiments have been conducted with traditional dimension reduction algorithms and clustering methods. The results of experiments demonstrate that, in most cases, the proposed scheme outperforms the traditional dimension reduction algorithms with different clustering methods.

1. Introduction

Anomaly detection is a very important branch of machine learning, with a wide range of practical applications, and it aims to detect special points in data. It is suitable for fault diagnosis [1, 2], system health monitoring [3], network security detection [4], intrusion and fraud detection [5–7], measurement, and other fields. The exceptions to the normal instances are called anomalies, so anomalies are also called exceptions, outliers, novelties, noises, and deviations [8]. The

so-called anomaly detection is to find objects that are different from most objects. The three objects O_1 , O_2 , and O_3 in Figure 1 are different from most of the objects in N_1 and N_2 classes. The deviation is different for different applications. Different application scenarios have different definitions of anomalies.

In order to find outliers in a given input sample, density estimation is a critical step. Although fruitful results have been achieved in unsupervised anomaly detection methods of data in recent years, there are always limitations on high-

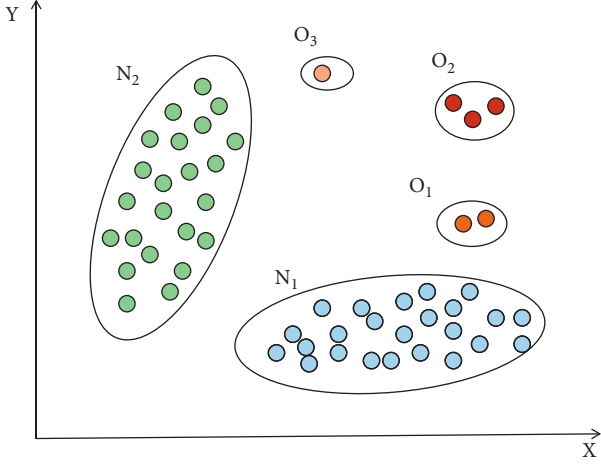


FIGURE 1: The data objects in N_1 and N_2 account for the majority of all data objects and are regarded as normal, while O_1 , O_2 , and O_3 are only a minority; they are regarded as abnormal.

dimensional data and multidimensional data. Some traditional dimension reduction methods, like Linear Discriminant Analysis (LDA), least absolute shrinkage and selection operator (LASSO), Locally Linear Embedding (LLE), Principal Component Analysis (PCA), Independent Principal Component Analysis (ICA), and Multidimensional Scale Transformation (MDS), are employed to process data, but, in the process of dimension reduction, some key information of the original data will be lost, which reduces the difference between normal samples and abnormal samples. In addition, there are some other methods, for example, clustering in the subspace of high-dimensional data [9], to further improve the anomaly detection results. However, none of the above methods can achieve the desired effect in the end. When deep neural networks have achieved good results in other fields, the dimensional disaster of data in anomaly detection seems to come to a turning point. Much research has actively been explored in this area. For example, the deep autoencoding Gaussian mixture model [10] has shown good performance on public datasets, providing a new direction for high-dimensional data anomaly detection.

According to the above analysis, we propose an anomaly detection scheme based on deep autoencoder. The following contributions are made to the unsupervised anomaly detection of high-dimensional data:

- (i) A dimension reduction method based on deep autoencoder and reconstruction of input samples is proposed. The deep autoencoder is used to reduce the dimension of the data, and the combination of the dimension reduction result and the reconstruction error forms a low-dimensional reconstruction input sample. The key information of the data is well preserved in the low-dimensional reconstruction input samples, which makes it easier to identify abnormal samples.
- (ii) An anomaly detection scheme based on deep autoencoder and clustering methods is proposed. This scheme makes full use of the advantages of the

deep autoencoder (DAE) to generate low-dimensional representation and reconstruction errors for the high-dimensional or multidimensional input data.

- (iii) A large number of comparison experiments have been conducted, and the experimental results on three public datasets prove that the scheme of a deep autoencoder-based and clustering method has better performance to identify abnormal points in the data.

2. Related Works

The probability of anomalies is small and complex, so it is difficult to identify all anomalies. Despite the emergence of many anomaly detection methods, the false alarm rate on the dataset [11] is still very high. In the anomaly detection of high-dimensional data, anomalies in the high-dimensional space are hidden, but they are obviously exposed in the low-dimensional space. Since the data generated in the real world is very complex, the anomaly detection of high-dimensional data is a difficult task [12]. Due to the large and complex data in the real world, it is also difficult to label the data. Generally, unsupervised methods [13] are used for anomaly detection. Usually, before the training and testing of anomaly detection, it is necessary to reduce dimension of high-dimensional data. Traditional dimension reduction methods include Linear Discriminant Analysis (LDA), least absolute shrinkage and selection operator (LASSO), Locally Linear Embedding (LLE), Principal Component Analysis (PCA), Linear Principal Component Analysis [14], and Nonlinear Principal Component Analysis [15, 16]. LDA is also called Fisher Linear Discriminant (FLD) because it was invented by Ronald Fisher in 1936. The basic idea of this method is to project samples of high-dimensional space into space of the best discriminant vector to achieve the purpose of extracting key information and compressing the dimension of Eigenvector-Space. After projection, it is ensured that the original sample has the largest interclass distance and the smallest intraclass distance in the new subspace; in other words, there is the best separability in the original sample. LASSO, which is a compression estimation method [17], obtains a more refined model by constructing a penalty function and achieves the purpose of dimensionality reduction by compressing some coefficients. When reducing dimensionality, LLE [18] focuses on maintaining the local linear features of the sample. Since LLE maintains the local features of the sample during dimensionality reduction, it is widely used in image recognition, visualization of high-dimensional data, and other fields. The main idea of PCA is to project n -dimensional data onto k -dimensional data to compress and denoise the data. The main features of the high-dimensional space data are projected to the low-dimensional space direction where the reconstruction error is minimized, and the most variables are retained as much as possible, so that the key information that distinguishes normal samples and abnormal samples in the original data is stored in the low-dimensional space. ICA is to separate independent signals from mixed observation signals or use

independent signals to represent other signals as much as possible. The idea of ICA was first proposed by Herault and Jutten in 1986, and it has been a powerful data analysis method in recent years. It is a method used to find hidden components from high-dimensional data and is considered an extension of PCA. The main idea of MDS is to map the coordinate points in the high-dimensional space to the low-dimensional space and keep the similarity between the data as much as possible. The main problem it solves is to give the similarity between m objects and determine the low-dimensional representation of the object so that it can match the original similarity to the greatest extent. In a high-dimensional space, a point represents an object, so the similarity between objects is related to the distance between points. The closer the distance between two points, the higher the similarity. Besides the dimension reduction method, the method based on subspace [9, 19] is also an alternative solution. In addition, recent dimension reduction and reconstruction errors based on deep autoencoding [10, 20] have made new progress. However, the process requires joint training of data dimensionality reduction, reconstruction error, and density estimation, which is much more complicated and requires a lot of time and computing resources.

The clustering methods are one important type of methods for anomaly detection and density estimation, including K-Means, Mean-Shift, DBSCAN, Gaussian mixture model, and multivariate mixture model [12, 21–26]. Due to the limitation of high-dimensional data [27], the above-mentioned methods cannot be directly applied to anomaly detection of high-dimensional data. For this problem, the traditional dimension reduction methods are generally used to preprocess the data. However, the key information of the sample data will be lost during the dimension reduction, which will cause the difficulty to identify anomalies in subsequent process. Recent studies have shown that the deep autoencoder method [10] that introduces reconstruction error could solve this problem well, because the deep autoencoder could eliminate less relevant features during the dimension reduction and retained the key information of the original data. According to the above analysis, in this paper, we propose an anomaly detection scheme based on deep autoencoder and clustering methods. The deep autoencoder can obtain low-dimensional data and reconstruction error, and both of them are further reconstructed to generate input samples, which gives full play to the advantages of deep autoencoder.

2.1. Dimension Reduction Method Based on Deep Autoencoding. The attributes of the data generated in the real world are not single, and the data of these multiple attributes forms a high-dimensional dataset. Because high-dimensional data not only occupies a huge storage capacity but also consumes much computing resources, it is imperative to reduce the dimension of high-dimensional data. The deep autoencoder can reconstruct the input through high-order features to achieve the purpose of dimensionality reduction. It is composed of two symmetrical neural networks, which are encoder and decoder.

2.2. Deep Autoencoder. The deep autoencoder is composed of two symmetrical, feedforward multilayer neural networks, namely, the encoder and the decoder, as shown in Figure 2. Here, the input data is fed to encoder for encoding. In the process, the Compressed Feature Vector is obtained, which is decoded by the decoder to receive the output data similar to the original space. In addition, the circular symbol represents the dimension of data in Figure 2. The data input dimension of the autoencoder is equal to the output dimension. With the help of sparse coding, a small number of high-order features are recombined to reconstruct the input instead of just copying pixels. It is usually used to learn the representation or coding of a set of input data, and its essence is to remove redundant information, so that the features of data are retained as far as possible.

An autoencoder is a neural network that reproduces the input signal. In order to reproduce the input data, the autoencoder must capture the most critical features that can represent the input data. When the number of intermediate hidden layer nodes is less than the number of input nodes, only the most important features in the data can be learned at this time. It can restore and remove redundant information. Similar to PCA, it looks for the main components that can represent the original data. In addition, regularization can be introduced in the intermediate hidden layer to penalize the sparseness of hidden layer nodes.

The desired output of the deep autoencoder is the input itself [21]. Let X be the input data sample; the encoder will map the input data sample X to the so-called latent representation Z_1 according to equation (1). Z_1 is fed to the decoder and it will be mapped to the output vector X' . X' is the corresponding expression of X , and it is usually impossible to completely reconstruct X . Therefore, there is error between them.

The expression of Z_1 is as follows:

$$Z_1 = \sigma(WX + b). \quad (1)$$

Here, Z is the latent representation, and σ denotes the activation functions. W denotes the weight, and b is the bias. The expression of X' is as follows:

$$X' = \sigma'(W'Z_1 + b'). \quad (2)$$

2.3. Reconstructing the Input Sample. The sources of reconstructed input sample composition are as follows:

- (1) The deep autoencoder reduces the dimension of the input sample X ; in the process, the potential representation Z_1 is obtained, as shown in equation (1).
- (2) Calculate the error between the input sample X and the output vector X' , which generates Z_2 in the process, as shown in the following equation:

$$Z_2 = f(X, X'). \quad (3)$$

Recombine Z_1 and Z_2 to form a low-dimensional input Z as follows:

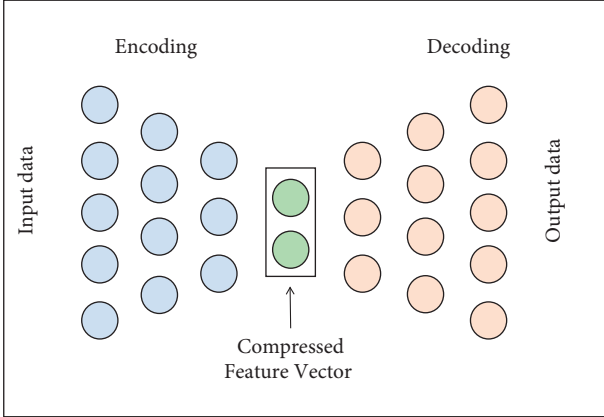


FIGURE 2: Structure of the deep autoencoder.

$$Z = [Z_1, Z_2], \quad (4)$$

where $f(\cdot)$ is the function of calculating reconstruction error. The dimension of Z_2 depends on the error obtained by several distance metrics, including absolute Euclidean distance, relative Euclidean distance, and cosine similarity [10].

2.4. Unsupervised Anomaly Detection Scheme. In this paper, a scheme is proposed for unsupervised anomaly detection, which is shown in Figure 3, using a reconstructed input network to obtain the compressed information, which is fed to clustering methods to identify the anomalies.

The main component of input networks is a deep autoencoder. Its purpose is to produce a low-dimensional representation of high-dimensional data, avoiding the limitation of data dimension on anomaly detection algorithms. As shown in Figure 4, the reconstructed input network works as follows: (1) It uses deep autoencoder to encode and decode data samples. (2) It reconstitutes low-dimensional input samples from the results and errors after reducing dimension of data.

In Figure 4, X is the input high-dimensional data, and Z_1 refers to the low-dimensional data compressed by the deep autoencoder. X' is obtained by the deep autoencoder decodes Z_1 ; and x' is similar to X . Z_2 is the reconstruction error obtained from x and X' . Z is the combination of Z_1 and Z_2 and is the final dimension reduction result. Moreover, the circular symbol represents the dimension of data in Figure 4.

In the proposed scheme, the clustering methods could be traditional clustering algorithms, like K-Means, DBSCAN, and Mean-Shift.

3. Experiment

In this section, to verify the effectiveness of the proposed scheme, massive comparison experiments have been conducted with traditional dimension reduction algorithms and clustering methods. The unsupervised anomaly detection methods to be verified include DAE + K-Means, DAE + DBSCAN, and DAE + Mean-

Shift. The deep autoencoder is trained to learn the compressed representation of the input data and then feed it to clustering approach, including K-Means, DBSCAN, and Mean-Shift. At the same time, we use traditional dimension reduction methods and deep autoencoder-based dimension reduction methods to conduct comparative experiments. These methods include Principal Component Analysis (PCA), Independent Principal Component Analysis (ICA), and multidimensional scaling transformation (MDS).

The experiment uses the following hardware configuration: MacBook Pro 2020, Intel Core i5 CPU, with 16 GB 2133 MHz LPDDR3 memory.

3.1. Dataset. We use several public datasets to conduct experiments to further observe the effects of unsupervised anomaly detection algorithms based on autoencoder and traditional unsupervised anomaly detection algorithms on different datasets.

The following will briefly introduce several public datasets used in the experiment:

- (i) **Thyroid:** Thyroid dataset is derived from the thyroid research cases of the Garavan Institute and can be obtained in the UCI machine learning warehouse. The dataset contains 15 categories and 6 real attributes. The data is divided into three categories: normal (not hypothyroidism), hyperfunction, and subnormal function based on whether the referred patient is hypothyroidism. In the original data, hyperfunction (hyperfunction) accounts for a small proportion of the total sample and is regarded as abnormal [10]. The other two categories that account for a larger proportion of the total sample are regarded as normal categories.
- (ii) **Arrhythmia:** Arrhythmia dataset comes from H. Altay Guvenir, Ph.D., and can be obtained in the UCI machine learning warehouse. The dataset is a multiclass classification dataset with a dimension of 279. Five category attributes were discarded in the experiment, so the total attribute is 274. The smallest sample category in the dataset [10], namely, 3, 4, 5, 7, 8, 9, 14, and 15, is combined into the outlier category, and the rest is merged into the normal category.
- (iii) **Pen_global:** Pen_global dataset was contributed by Markus Goldstein to the Dataverse project on October 6, 2015, and can be used for unsupervised tasks. This project is dedicated to helping researchers access and use data. The Pen_global dataset has a total of 17 attributes.

The details of each dataset are shown in Table 1.

3.2. Clustering Methods. We use traditional clustering algorithms to perform anomaly detection on reconstructed input samples, including K-Means, DBSCAN, and Mean-Shift.

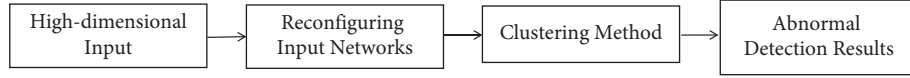


FIGURE 3: A framework for unsupervised anomaly detection methods.

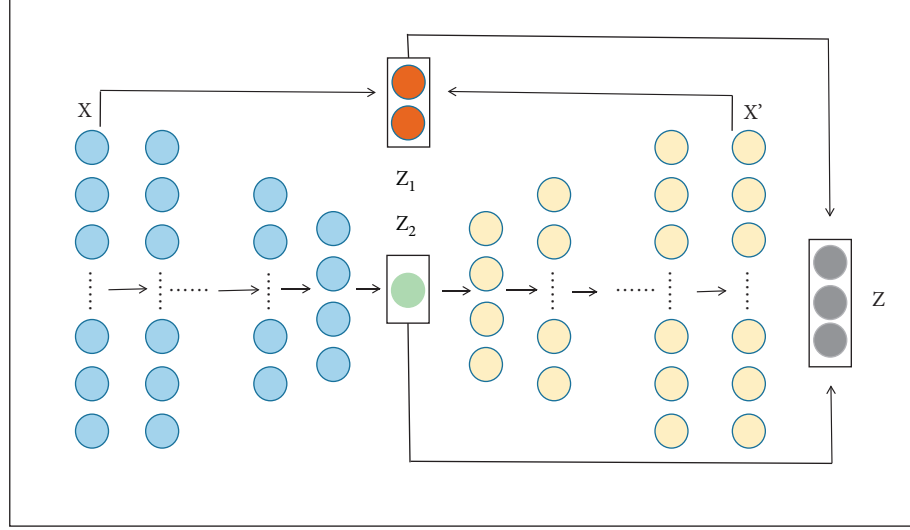


FIGURE 4: Reconfiguring input networks.

TABLE 1: Details of the datasets.

Methods	Dimensions	Instances	Anomaly ratio (ρ)
Thyroid	6	3,772	0.025
Arrhythmia	274	452	0.15
Pen_global	16	809	0.11

3.2.1. K-Means. The main idea of the K-Means is as follows: first, initialize k points as the center of each cluster, and divide the data points close to the cluster center into a cluster. The data is divided for the first time to obtain k clusters. Recalculate the Euclidean distance from each data point to the cluster center, and take the average to update the center point of each cluster. Redivide the data points so that they are the closest to the cluster center. Until the cluster center does not change, the iteration stops.

K-Means is a distance-based clustering algorithm, which aims to cluster similar samples into one category, so that samples of different categories are as far away as possible [28], so samples of different categories can be separated. When there are only two types of samples, one is called normal and the other is abnormal. There are two situations that can be considered abnormal. One case is that a sample is very close to the center of the abnormal class relative to the center of the normal class; the other case is if the distance between a sample and the center of the normal class is greater than a pre-determined threshold [21]. The sample points P_1 and P_2 in Figure 5 correspond to the above two situations, respectively.

3.2.2. DBSCAN. Given m data, if the ϵ neighborhood of an object contains at least m objects, the object is called the core object. Given a set of objects D , if p is in the ϵ neighborhood

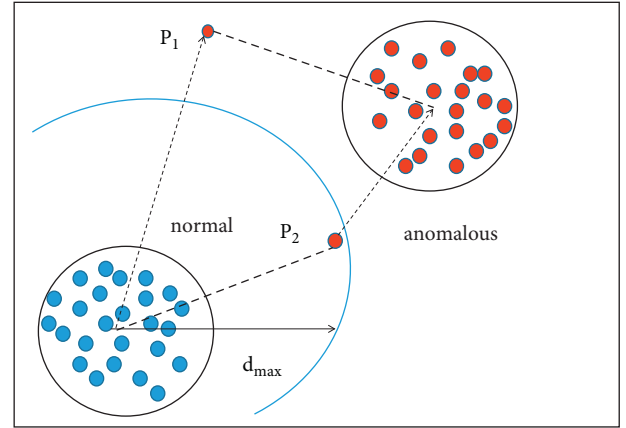


FIGURE 5: Outlier detection.

of q and q is a core object, then object p is directly density-reachable from object q , as shown in Figure 6. If there is an object chain $p_1 p_2 \dots p_m$, $p_1 = q$, $p_n = p$, for p_i belongs to D , p_{i+1} is directly reachable from p_i on ϵ and m , and then p is density-reachable from q on ϵ and m .

The main idea of the DBSCAN is to randomly select a core object without a category as a seed and use all the density-reachable sample sets of the core object as a cluster. Then select another core object with no category to find a set of samples with reachable density to obtain another cluster. Run iteratively until all core objects have categories.

DBSCAN is a density-based clustering method [29], which aims to find categories of arbitrary shapes in data. In this algorithm, the category can be regarded as the sample dense area divided by the sample low-density area in the data space [22]. Therefore, it can be used to detect anomalies in data samples.

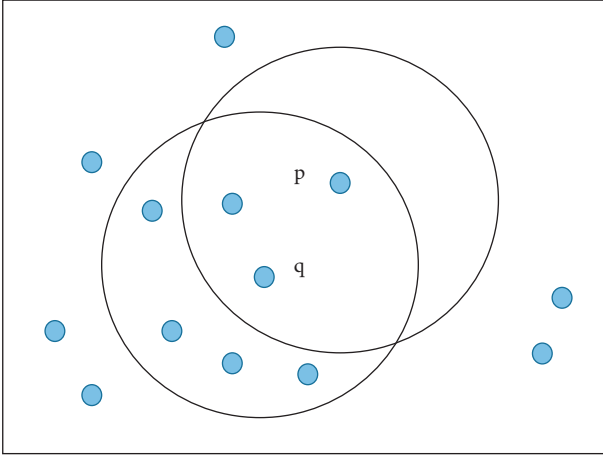


FIGURE 6: Directly density-reachable.

3.2.3. Mean-Shift. The main idea of Mean-Shift is to calculate the average value M of the vector distance between a certain point P and its surrounding radius R and calculate the direction in which the point will drift (move) in the next step. When the point no longer changes, it forms a cluster with the surrounding points, calculate the distance between the cluster and the historical cluster; if it meets the condition of less than threshold D , it can be merged into the same cluster; otherwise, it forms itself a cluster, until the selection of all data points is completed.

Mean-Shift is also a density-based clustering algorithm [30, 31]. The algorithm updates the centroid to the average value of the specified area through iteration to achieve the purpose of clustering [23, 24]. Due to the distance between the sample and the offset point being different, the contribution of the corresponding offset to the mean offset vector is also different. Therefore, in order to solve the problem, it needs to find density function by introducing the kernel function.

4. Results

In this part, we use precision and F1 score to evaluate the anomaly detection performance of the algorithm. Table 2 shows the precision, F1 score, and time of the experimental results on Thyroid dataset. Table 3 shows the precision, F1 score, and time of the experimental results on Arrhythmia dataset. Table 4 shows the precision, F1 score, and time of the experimental results on Pen_global dataset. The best results of each algorithm are shown in bold.

Table 2 shows the precision, F1 score, and time index values of each algorithm on the Thyroid dataset. It can be found that DAE + K-Means, DAE + DBSCAN, and DAE + Mean-Shift achieve the best results in precision and F1 score indicators. Table 3 shows the precision, F1 score, and time index values of each algorithm on the Arrhythmia dataset. It can be found that DAE + K-Means, DAE + DBSCAN, and DAE + Mean-Shift achieve the best results in precision and F1 score indicators. Table 4 shows the precision and F1 score and time index values of each algorithm on the Pen_global dataset. DAE + K-Means,

TABLE 2: Experimental results on Thyroid dataset.

	Thyroid		
	Precision	F1	Time (s)
K-Means	0.0246	0.0471	0.8842
PCA + K-Means	0.0390	0.0727	0.6798
ICA + K-Means	0.0267	0.0514	0.2449
MDS + K-Means	0.0217	0.0417	17.7753
DAE + K-Means	0.2537	0.3981	0.9136
DBSCAN	0.2432	0.3803	15.1509
PCA + DBSCAN	0.1340	0.2342	0.62020
ICA + DBSCAN	0.1833	0.3034	8.4805
MDS + DBSCAN	0.1625	0.2549	24.5062
DAE + DBSCAN	0.8000	0.5594	3.4793
Mean-Shift	0.3077	0.4571	1.2240
PCA + Mean-Shift	0.2609	0.3810	1.4040
ICA + Mean-Shift	0.3016	0.4419	0.9795
MDS + Mean-Shift	0.3023	0.4333	18.4106
DAE + Mean-Shift	0.8529	0.7205	1.5510

TABLE 3: Experimental results on Arrhythmia dataset.

	Arrhythmia		
	Precision	F1	Time (s)
K-Means	0.1899	0.2776	0.7734
PCA + K-Means	0.1538	0.2424	0.6006
ICA + K-Means	0.1132	0.1846	0.1842
MDS + K-Means	0.1333	0.2143	0.3223
DAE + K-Means	0.3545	0.4432	0.7923
DBSCAN	0.2457	0.3772	0.0537
PCA + DBSCAN	0.2000	0.3056	1.0822
ICA + DBSCAN	0.1400	0.2295	0.6780
MDS + DBSCAN	0.1149	0.2000	0.8024
DAE + DBSCAN	0.5000	0.5217	0.5873
Mean-Shift	0.5000	0.4375	0.2449
PCA + Mean-Shift	0.5000	0.4615	0.8521
ICA + Mean-Shift	0.6667	0.4000	0.3851
MDS + Mean-Shift	0.5833	0.5000	0.5987
DAE + Mean-Shift	0.8158	0.5962	0.2980

TABLE 4: Experimental results on Pen_global dataset.

	Pen_global		
	Precision	F1	Time (s)
K-Means	0.8780	0.6261	0.0680
PCA + K-Means	0.8660	0.6914	0.5920
ICA + K-Means	0.8050	0.6860	0.1326
MDS + K-Means	0.9333	0.6335	0.8024
DAE + K-Means	0.9849	0.7716	0.7817
DBSCAN	0.8799	0.9315	0.0256
PCA + DBSCAN	0.8832	0.8612	1.4397
ICA + DBSCAN	0.8917	0.9333	1.0609
MDS + DBSCAN	0.1373	0.2240	1.7177
DAE + DBSCAN	0.9420	0.9675	0.44910
Mean-Shift	0.7867	0.5339	0.0855
PCA + Mean-Shift	0.8478	0.6555	0.5777
ICA + Mean-Shift	0.6667	0.4000	0.2387
MDS + Mean-Shift	0.7778	0.6471	0.9347
DAE + Mean-Shift	0.9265	0.8097	0.0865

DAE + DBSCAN, and DAE + Mean-Shift achieve the best results in precision and F1 score indicators.

According to the experimental results, there is a significant improvement in the performance of identifying anomalies by using DAE. Although DAE + K-Means, DAE + DBSCAN, and DAE + Mean-Shift do not achieve the best result in the index of time, make the best result in the index of precision and F1 score. For instance, on the index of time, the value of ICA + Mean-Shift is 0.9795, while the value of DAE + Mean-Shift is 1.5510. The difference between the two is only 0.5715. Therefore, it is worth spending a few time to obtain better algorithm performance. On the whole, compared with other dimension reduction methods, including Principal Component Analysis (PCA), Independent Principal Component Analysis (ICA), and multidimensional scaling (MDS), the clustering algorithm with DAE has the best performance in anomaly detection. Compared with PCA, ICA, and MDS, the difference of dimension reduction based on DAE is the composition of compressed information. The former is to eliminate redundant information in data, which may lose important information in the original data, while the latter is to add reconstruction error based on the former. The compressed information obtained by DAE preserves the key of information in original data, which is critical for identifying anomalies.

5. Conclusion

Due to the complex reality scene, the generated data has the characteristics of large volume and high dimension. It can be seen that not all data can be used directly, and anomaly detection is often limited by the dimensionality of the data. The best way to solve this problem is to reduce the dimensionality of the data before detecting data anomalies. In this manuscript, we analyze the limitations of existing dimensionality reduction techniques and propose solutions to solve them. We proposed an unsupervised anomaly detection scheme based on DAE and clustering algorithms which can model the data efficiently. The clustering algorithms used in the experiment are K-Means, DBSCAN, and Mean-Shift. Experimental results show that our proposed scheme is effective in detecting anomalies in public datasets. In future work, we plan to apply the proposed unsupervised anomaly detection to network security data. Since there may be multiple anomalies in network security data, we plan to extend the binary classification problem to a multiclass classification problem. It can identify different types of abnormalities and improve the security performance of the network in the future.

Data Availability

The data used to support the findings of this study are 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 study.

Acknowledgments

This research was supported by Tianjin Municipal Science and Technology Bureau under Grant no. 18JCZDJC32100. The author Chuanlei Zhang received this grant and the URL of the sponsor's website is <http://kxjs.tj.gov.cn/>. This research was also funded by National Natural Science Foundation of China under Grants no. 51874300 and no. U1510115. The author Wei Chen received these grants and the URL of the sponsor's website is <http://www.nsf.gov.cn/>. This research was also funded by Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, under Grant no. 20190902. The author Wei Chen received the grant and the URL of the sponsor's website is <http://www.sim.ac.cn/>.

References

- [1] A. Purarjomandlangrudi, A. H. Ghapanchi, and M. Esmalifalak, "A data mining approach for fault diagnosis: an application of anomaly detection algorithm," *Measurement*, vol. 55, pp. 737-738, 2014.
- [2] C. Li, Y. Yang, and K. Zhang, "A fast MPPT-based anomaly detection and accurate fault diagnosis technique for PV arrays," *Energy Conversion and Management*, vol. 234, Article ID 113950, 2021.
- [3] R. A. Martin, M. Schwabacher, and N. Oza, "Comparison of unsupervised anomaly detection methods for systems health management using space shuttle," in *Proceedings of the Joint Army Navy NASA Air Force Conference on Propulsion*, 2007.
- [4] S. Guo, J. Zhao, and X. Li, "A black-box attack method against machine-learning-based anomaly network flow detection models," *Security and Communication Networks*, vol. 2021, pp. 1-13, 2021.
- [5] D. Massa and V. Raul, "A fraud detection system based on anomaly intrusion detection systems for e-commerce applications," *Computer and Information Science*, vol. 7, no. 2, pp. 117-140, 2014.
- [6] K. M. Al-Gethami, M. T. Al-Akhras, and M. Alawairdhi, "Empirical evaluation of noise influence on supervised machine learning algorithms using intrusion detection datasets," *Security and Communication Networks*, vol. 2021, pp. 1-28, 2021.
- [7] G. Lu and X. Tia, "An efficient communication intrusion detection scheme in ami combining feature dimensionality reduction and improved LSTM," *Security and Communication Networks*, vol. 2021, pp. 1-21, 2021.
- [8] L. Portnoy, *Intrusion Detection with Unlabeled Data Using Clustering*, Columbia University, New York, US, 2001.
- [9] P. Lance, E. Haque, and L. Huan, "Subspace clustering for high dimensional data: a review," *ACM sigkdd explorations newsletter*, vol. 6, pp. 90-105, 2004.
- [10] B. Zong, Q. Song, and M. Min, "Deep autoencoding Gaussian mixture model for unsupervised anomaly detection," in *Proceedings of the International Conference on Learning Representations*, Vancouver, BC, Canada, 2018.
- [11] G. Pang, C. Shen, and A. van den Hengel, "Deep anomaly detection with deviation networks," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 353-362, Anchorage, AK, USA, 2019.
- [12] A. Zimek, E. Schubert, and H.-P. Kriegel, "A survey on unsupervised outlier detection in high-dimensional numerical

- data,” *Statistical Analysis and Data Mining*, vol. 5, no. 5, pp. 363–387, 2012.
- [13] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation-based anomaly detection,” *ACM Transactions on Knowledge Discovery from Data*, vol. 6, no. 1, pp. 1–39, 2012.
 - [14] I. T. Jolliffe, “Principal components in regression analysis,” *Principal Component Analysis*, pp. 129–155, 1986.
 - [15] S. Günter, N. Schraudolph, and S. Vishwanathan, “Fast Iterative Kernel Principal Component Analysis,” *The Journal of Machine Learning Research*, vol. 8, pp. 1893–1918, 2007.
 - [16] E. Ronchetti, “Robust inference,” *International Encyclopedia of Statistical Science*, Springer, Berlin/Heidelberg, Germany, pp. 1240–1242, 2011.
 - [17] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society: Series B*, vol. 58, no. 1, pp. 267–288, 1996.
 - [18] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
 - [19] T. Pevný, “Loda: lightweight on-line detector of anomalies,” *Machine Learning*, vol. 102, no. 2, pp. 275–304, 2016.
 - [20] C. Zhou and R. C. Paffenroth, “Anomaly detection with robust deep autoencoders,” in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 665–674, Halifax, NS, Canada, August 2017.
 - [21] G. Münz, S. Li, and G. Carle, *Traffic Anomaly Detection Using K-Means Clustering*, pp. 13–14, GI/ITG Workshop MMBnet, 2007.
 - [22] Z. Chen and Y. F. Li, “Anomaly detection based on enhanced DBScan algorithm,” *Procedia Engineering*, vol. 15, pp. 178–182, 2001.
 - [23] M. Netzer, J. Michelberger, and J. Fleischer, “Intelligent anomaly detection of machine tools based on mean shift clustering,” *Procedia CIRP*, vol. 93, pp. 1448–1453, 2020.
 - [24] D. Comaniciu and P. Meer, “Mean shift: a robust approach toward feature space analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
 - [25] L. Xiong, B. Póczos, and J. Schneider, “Group anomaly detection using flexible genre models,” in *Proceedings of the 24th International Conference on Neural Information Processing Systems*, Granada, Spain, December 2011.
 - [26] J. Kim and C. D. Scott, “Robust kernel density estimation,” *Journal of Machine Learning Research*, vol. 13, pp. 2529–2565, 2012.
 - [27] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, “High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning,” *Pattern Recognition*, vol. 58, pp. 121–134, 2016.
 - [28] M. Tahmid Rahman Laskar, J. Huang, and V. Smetana, “Extending Isolation Forest for Anomaly Detection in Big Data via K-Means,” arXiv e-prints: <http://arxiv.org/abs/2104.13190>, 2021.
 - [29] Y. Chen, S. Tang, N. Bouguila, C. Wang, J. Du, and H. Li, “A fast clustering algorithm based on pruning unnecessary distance computations in DBSCAN for high-dimensional data,” *Pattern Recognition*, vol. 83, pp. 375–387, 2018.
 - [30] M. A. Carreira-Perpinán, “A review of mean-shift algorithms for clustering,” arXiv preprint <http://arxiv.org/abs/1503.00687>, 2015.
 - [31] J. Jordan and E. Angelopoulou, “Mean-shift Clustering for Interactive Multispectral Image Analysis,” in *Proceedings of the 2013 IEEE International Conference on Image Processing*, pp. 3790–3794, Melbourne, Victoria, September 2013.