



Deep Learning Based Anomaly Detection for Multi-dimensional Time Series: A Survey

Zhipeng Chen¹, Zhang Peng^{2,3}, Xueqiang Zou¹(✉), and Haoqi Sun^{2,3}

¹ National Internet Emergency Center, CNCERT/CC, Beijing 100029, China
zouxueqiang@cert.org.cn

² School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100864, China

³ Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

Abstract. Multi-dimensional time series are multiple sets of variables collected in chronological order, which are the results of observing a certain potential process according to a given sampling rate. It also has the ability to describe space and time and is widely used in many fields such as system state anomaly detection. However, multi-dimensional time series have problems such as dimensional explosion and data sparseness, as well as complex pattern features such as periods and trends. Such characteristics lead to rule-based anomaly detection methods suffer from poor detection effects. In the big data scenario, deep learning method begins to be applied to anomaly detection tasks for multi-dimensional time series due to its wide coverage and strong learning ability. This work first summarizes the definitions of anomaly detection for multi-dimensional time series and the challenges it faces. Related methods are sorted out, and then the deep learning-based method is emphasized. The existing work and its advantages and disadvantages are summarized. Finally, the shortcomings of the existing algorithms are clarified and the future research direction is explored.

Keywords: Anomaly detection · Muti-dimensional time series · Machine learning · Deep learning

1 Introduction

With the rapid development of the information age, the amount of data has exponentially increased. In the environment of big data, how to mine the hidden information from the massive data is a new topic and challenge for the development of information technology. At the same time, there are a lot of abnormal patterns in these data, which also contain a lot of important information. Anomaly detection actually refers to finding the data that does not match the normal pattern in the data [1]. Such mismatched data is called anomaly or outlier. Abnormal patterns and outliers are two types of detected entities in anomaly detection task. As shown in Fig. 1, the types of abnormalities mainly include abnormal points and abnormal sequences. In anomaly detection, traditional heuristic rules set thresholds based on historical data. If the value of a point exceeds or falls below the threshold, the point is classified as a “point anomaly”. At present, most of the existing work is aimed at the detection of abnormal points.

© The Author(s) 2022

W. Lu et al. (Eds.): CNCERT 2021, CCIS 1506, pp. 71–92, 2022.

https://doi.org/10.1007/978-981-16-9229-1_5

An abnormal sequence refers to a continuous abnormal pattern in data points within a continuous period of time. For example, a time series has the same trend every day from 7 am to 9 am, indicating that the series has a certain periodicity. If the trend changes on a certain day, it is more likely that there will be an anomaly on that day. The anomaly detection method for sequence is more complex than anomaly points, and the real-time effect of the algorithm is poor.

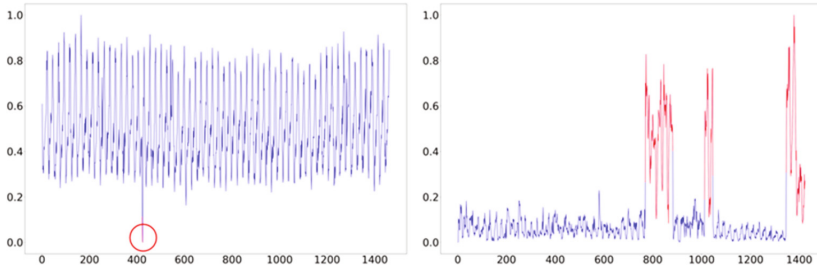


Fig. 1. Two types of anomalies (points and sequence).

An abnormal sequence refers to a continuous abnormal pattern in data points within a continuous period of time. For example, a time series has the same trend every day from 7 am to 9 am, indicating that the series has a certain periodicity. If the trend changes on a certain day, it is more likely that there will be an anomaly on that day. The anomaly detection method for sequence is more complex than anomaly points, and the real-time effect of the algorithm is poor.

There are many reasons for the abnormal pattern of the data. Firstly, there may be an error in the system, causing some noise and missing values in the data; secondly, there may be that an unknown or deviated data from the normal pattern is generated in the system, which means that there is an abnormality. When analyzing the real data set in the real world, it is necessary to identify the abnormal pattern, that is, to distinguish the data points in the data sample set that deviate from the normal pattern, and to dig out the relevant information hidden in the abnormal data. At present, anomaly detection is applied in many fields, such as fraud detection, network intrusion detection, medical anomaly detection, log anomaly detection, video surveillance anomaly detection, and industrial IoT big data anomaly detection. At the same time, the data entities in most of the above fields are typical time series data.

Time series data are recorded periodically in the form of series for the data describing the system behavior at each moment. System behavior may change due to some external events or changes in the internal state of the system. Therefore, for time series data, a large amount of system-related information can be mined from factors such as data trends, peaks, valleys, and periodicity.

Multi-dimensional time series are groups of ordered variables collected according to time sequence and a given sampling frequency. They are the result of observing a certain potential process and have the ability to describe space and time at the same

time. Broadly speaking, a multi-dimensional time series is composed of multiple single-dimensional time series. For example, the CPU utilization rate collected from a server is a single-dimensional time series, while a multi-dimensional time series records multiple system indicators at the same time. In the system operation and maintenance monitoring scenario, the monitoring of the real-time status of the database includes multiple indicators such as the number of transactions per second, the number of active sessions, and the number of connected sessions and other indicators. That is, the real-time status of the database will be determined by multiple indicators.

Time series data is strongly correlated with time. The longer the time period, the more data generated, and the greater the amount of time series data in the time dimension. Existing work mostly uses sliding windows to cut the time series and divide them into multiple sub-sequences with smaller dimensions before analysis. As shown in Fig. 2, assuming that the window size of the sliding window is 4 and the step size is 2, a sequence with a length of 12 is divided into 5 sub-sequences, and there is overlap between the sub-sequences. The window and step length parameters should be analyzed in detail according to the specific algorithm.

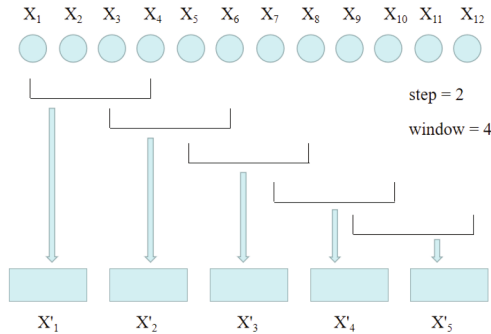


Fig. 2. Use sliding window to segment spatio-temporal series.

In addition, the multi-dimensional time series contains complex periodic components, trend components, and high-frequency residual components. As shown in Fig. 3, concept drift, periodic change and other phenomena may occur in different time periods of a time series, and there is certain noise. The particularity of multi-dimensional sequence data makes it difficult to directly model time series and apply them to downstream anomaly detection tasks.

In this context, methods such as heuristic rules based on the establishment of alarm thresholds are no longer suitable for anomaly detection in big data scenarios. The industry expects to learn the internal correlation and essence of massive data through the idea of deep learning, and detect the possible anomalies in the system through artificial intelligence. At present, how to apply deep learning and other algorithms to anomaly detection tasks oriented to multi-dimensional time series is also one of the research hotspots in the industry and academia in recent years.

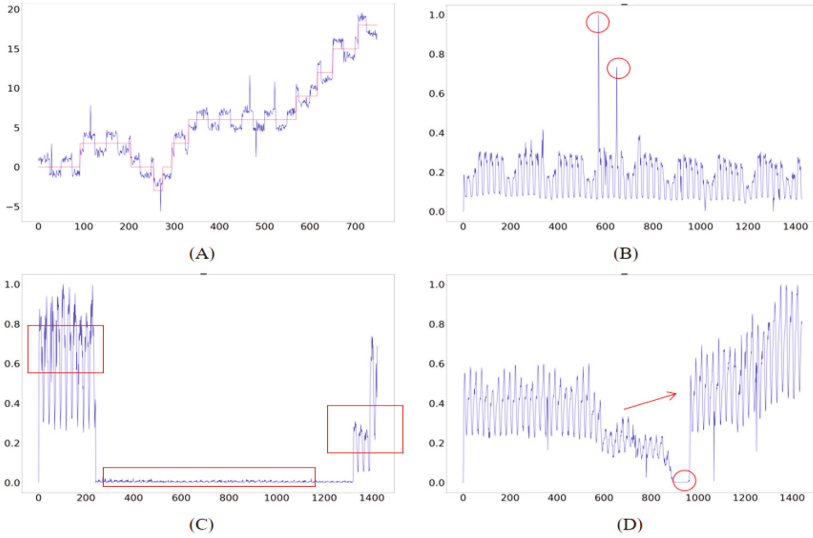


Fig. 3. Complex pattern in time series.

In this paper, the multi-dimensional time series-oriented anomaly detection technology and methods are described and summarized. The main arrangements of the paper are as follows: the first part of the introduction mainly introduces related definitions and research background, and the second part summarizes the challenges and difficulties faced by the multi-dimensional time series-oriented anomaly detection task. The third to fifth parts mainly organize and analyze different anomaly detection methods. Which focuses on summarizing the anomaly detection methods based on deep learning. Finally, the sixth part summarizes the shortcomings of anomaly detection methods based on deep learning, and looks forward to the future research direction.

2 Challenge

Multi-dimensional time series data is strongly related to time. The data at a certain moment records the real-time status information of the system at that moment, which has the characteristics of dimension explosion and data imbalance, and most of the application fields require high real-time performance for anomaly detection. At the same time, the time series contains a large number of complex temporal and spatial semantic features such as cycles and trends, which brings challenges to anomaly detection tasks for multi-dimensional time series.

2.1 Dimensional Explosion

The time series is strongly correlated with time. The longer the time period, the higher the time dimension of the collected data. Since the monitoring of the system is $7 * 24$ uninterrupted, the amount of monitoring data will continue to grow as time goes by. At

the same time, in the process of collecting data, some data may be lost due to sensor failure and other reasons; noise data may also be collected due to system failure and other reasons. For this kind of data, it is necessary to preprocess the noise and missing values in the time series to reduce the dimensionality of the data before performing anomaly detection.

2.2 Concept Drift

In the real world, time series are generally a non-stationary series, that is, the mean and variance of the time series do not obey a certain distribution, so there are very large limitations in the feature representation. For example, the difference integrated moving average autoregressive model (ARIMA) [2] has poor prediction effect for non-stationary series. In the process of time series anomaly detection, as time changes, both the abnormal mode and the normal mode of the sequence may change, that is, the phenomenon of concept drift occurs. At the same time, change points may also occur in the time series. These factors will affect the accuracy of anomaly detection results.

2.3 Complex Semantics

Time series data has natural time semantics, that is, the system state at time $t + 1$ may be related to the system state from time 1 to t . In addition, time series may have characteristics such as periodicity and periodicity. For example, a company draws and analyzes the historical data of system monitoring indicators, and finds that the data line chart shows some similar trends at a specific time of each day, from which some system related information can be digged out.

There are not only temporal semantics but also spatial semantics in multi-dimensional time series data. The concept of spatial semantics is mainly derived from the spatial dimensions of multi-dimensional time series data. When monitoring the state of a system, there may be multiple monitoring indicators, and there is a certain correlation between data of different dimensions and jointly determine the current system status. Therefore, for multi-dimensional time series data, it is necessary to mine spatial semantics in spatial dimensions.

When analyzing time series data, there may also be external semantic features in the data. For example, the UAH-DriveSet [3] dataset collects driving behavior data during six driving sessions, and records the speed and direction of the car at each moment. At the same time, the data set additionally records the type of road (highway, urban road, village lane, etc.) that is driven each time. The characteristics of this dimension have nothing to do with temporal and spatial semantics, but it has a key impact on the accuracy of anomaly detection results.

It can be seen that there are temporal semantics, spatial semantics and external semantics in multi-dimensional time series. The heuristic rule method based on threshold setting cannot mine rich and complex semantic features, which further affects the accuracy of anomaly detection.

2.4 Data Sparse

In a period of time, abnormal patterns only account for a small part, and most of the sequences are normal patterns. Similarly, in most public data sets related to time series, the number of abnormal patterns in the sample set is very small, and the imbalanced data sets cause certain classifiers in machine learning to have a certain bias. At the same time, there are very few public data sets for time series. The public data sets available are Yahoo Benchmark [4] and Numenta Anomaly Benchmark [5]. Since deep learning methods require a large amount of training data, the data volume of existing public data sets cannot meet the requirements. Existing works mostly use private data sets to expand the data volume, and the private data sets are generally unlabeled. On the one hand, because anomaly detection is a typical two-class or multi-classification task, it is difficult to use unsupervised learning algorithm directly because of its high dependence on data labels; on the other hand, if the algorithm adopts supervised and semi-supervised learning methods, it is necessary to label time series data, which requires strong professional knowledge and it is very labor intensive.

2.5 Poor Scalability

Anomaly detection is roughly divided into offline detection and online detection. Offline detection refers to the analysis of historical data to extract abnormal patterns; while online detection is the real-time analysis and monitoring of the system status. In an industrial environment, the version of the system will continue to change with the update of requirements and the improvement of the architecture, resulting in frequent changes in the data entities in the online inspection process, and the dimensionality of the collected time series data continues to increase. However, existing algorithms generally only analyze historical data, and the model has poor scalability, so it can not be applied to industrial production environments. In addition, the online detection method needs to control the calculation time delay of the algorithm to a lower range. If the abnormality can be detected earlier, the more the loss caused by the abnormality of the system can be recovered, but this puts higher requirements on the calculation time of the algorithm.

2.6 Summary

The challenges of anomaly detection algorithms for multi-dimensional time series are summarized in Table 1. The anomaly detection task is generally divided into four steps: data collection, data preprocessing, feature representation learning, and anomaly detection. In the above process, there may be problems such as data sparseness, noise and missing values, complex semantic information, and poor real-time performance of algorithms.

Table 1. Summary of challenges during multi-dimensional time series-oriented anomaly detection process.

Process	Challenge	Description
Data source	Lack of public dataset	Data points in existing public dataset [4, 5] are far from enough
	Unbalanced dataset	The proportion of abnormal points too low causes the Classifier to be biased
Data preprocess	Noise and missing value	Noise and missing value in real world data interfere with model performance
Feature learning	Complex contextual information	Temporal-spatial and external semantics are included in real world data which makes it too hard to extract them all
	Seasonal shift	Real world data contains seasonal component
	Trend change	Real world data contains trend component
	Concept drift	Real world data suffer from concept drift which makes the performance of trained model drop
Anomaly detection	High computational cost	Online mode calls for low latency which makes many existing expensive computational cost model not applicable

In recent years, there have been more related works based on heuristic rule methods to detect abnormal patterns in single-dimensional time series, and there have also been more mature applications in industrial system monitoring. However, with the increasing complexity of the system architecture, the anomaly detection entities have evolved from a simple single-dimensional time series to a multi-dimensional time series, resulting in a significant decline in the detection accuracy based on heuristic rule methods. In view of the above difficulties, the following chapters will sort out and analyze the rule-based anomaly detection methods, and discuss their limitations and deficiencies in detail.

3 Rule-Based Anomaly Detection Algorithm

The method based on heuristic rules has been applied more maturely in the task of anomaly detection for single-dimensional time series, and the detection effect is better. However, due to the increasing complexity of the system and the explosive growth of data volume, anomaly detection entities have evolved from a single-dimensional time series to a multi-dimensional time series, resulting in rule-based methods no longer suitable

for anomaly detection in a big data environment. This section will sort out the rule-based method and summarize its shortcomings.

The method based on heuristic rules is very simple and intuitive. By observing historical data, a maximum threshold and a minimum threshold are set manually. Once the value of a certain point exceeds a given range, it will be judged as an abnormal point. However, setting the threshold requires very strong prior knowledge and a large amount of historical data, which will consume a lot of manpower and material resources. At the same time, there may be phenomena such as conceptual drift in the time series. For example, when the system is upgraded, the distribution of the whole time series data will change, so that the previously set threshold may no longer be applicable, and the method's universality is poor.

An improvement to the heuristic rule is to introduce the concept of statistics, that is to calculate the mean and variance according to the historical data, and set the threshold automatically according to these indicators. Another similar statistical method is the box plot method. The box plot method divides the data into several "boxes" through the minimum non abnormal observation, lower quartile Q1, median, upper quartile Q3 and maximum non abnormal observation, and any data not in the box is classified as abnormal [6]. The box plot method is often used in the detection of abnormal points in the medical field.

The advantage of this type of method is that the algorithm has high real-time performance and can meet the requirements of real-time detection in terms of computing speed. It is suitable for the concept of real-time monitoring and alarm generation of machines and equipment in an industrial environment. However, the shortcomings of statistical methods lie in the inability to capture the spatial and temporal semantic characteristics of time series data, and the time series data in the real world are generally non-stationary, with periodicity, concept drift and other phenomena. There are some limitations in using the method of fitting distribution to divide the data. At the same time, for multi-dimensional time series, it is necessary to set a threshold for each dimension separately, which leads to the reduction of the usability and universality of the heuristic rule method. Therefore, the false negative rate and false positive rate are relatively high in the process of anomaly detection.

In recent years, machine learning algorithms have developed rapidly, and their theories and methods have been widely used to solve complex problems in engineering applications and scientific fields. Machine learning methods have good interpretability and strong generalization ability, and they are also widely used in anomaly detection tasks. The following will sort out and summarize the related work using machine learning methods.

4 Anomaly Detection Algorithm Based on Machine Learning

As a research hotspot in the field of pattern recognition and artificial intelligence, machine learning has been used to solve some complex problems in the industry and academia, including anomaly detection. In recent years, Yahoo has developed a time series anomaly detection framework EGADS [7], which belongs to the state-of-art

method in KPI anomaly detection. Broadly speaking, anomaly detection can be divided into three major categories, namely supervised methods, semi-supervised methods and unsupervised methods. The supervised methods require data sets to be labeled. However, most data sets in the industry are unlabelled, because data labeling consumes a lot of manpower and material resources. Therefore, it is relatively difficult to implement supervised methods. In recent years, some work tends to use semi supervised or unsupervised methods to detect anomalies in time series.

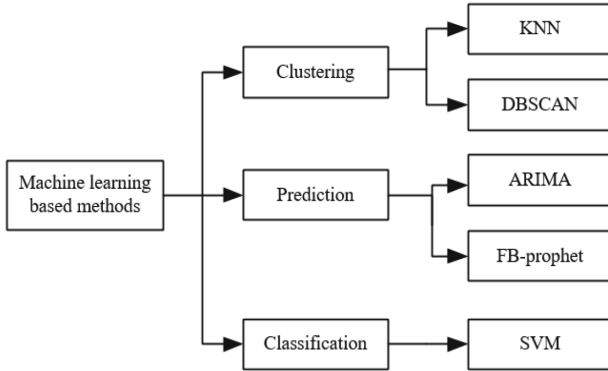


Fig. 4. Anomaly detection algorithms based on machine learning.

According to the principle of the method, the anomaly detection method based on machine learning can be divided into three parts: clustering based method, classification-based method and prediction method, as shown in Fig. 4. These three parts will be introduced in detail below.

4.1 Clustering-Based Method

Clustering is an unsupervised machine learning algorithm, which has a wide range of applications in the engineering field because the clustering algorithm does not require the data set to be labeled. The algorithm uses the idea of the distribution difference between normal points and abnormal points in the vector space to project them into the vector space. At present, some mainstream clustering algorithms mainly include the K-means algorithm (K-means), the nearest neighbor algorithm (KNN), the density-based clustering algorithm (DBSCAN), and the maximum expectation (EM) clustering using Gaussian Mixture Model (GMM).

Among them, Ramaswamy et al. [8] used the KNN algorithm [9] to detect anomalies in the data using a distance-based method, calculating the K proximity distance for each point in the data set. Then using a threshold method, once the distance exceeds the threshold, the point is judged to be abnormal. However, this algorithm requires manual determination of some parameters and abnormal thresholds, and is very sensitive to data changes. Li et al. [10] proposed a KPI clustering framework ROCKA to solve the

problem of too many training models caused by too many KPIs in industry. The framework first preprocessed KPIs and extracted KPI baselines, then using density aggregation DBSCAN [11] to cluster KPIs, and divides similar KPIs into the same category. According to the above ideas, bu et al. [12] extracted 14 dimensional features such as SVD [13], Holt winters [14], wavelet [15] for each KPI baseline to train anomaly detection model, which greatly reduced the number of models to be trained. In general, clustering methods are widely used in the field of anomaly detection. However, because clustering methods divide data points by distance, density or distribution, they still cannot capture the temporal and spatial semantics of time series.

4.2 Classification-Based Method

The classification-based method uses the given data label in the training set or a custom anomaly threshold to train the model and classify the data. At present, the commonly used classification algorithms in the field of anomaly detection include support vector machine SVM [16], isolated forest and random forest.

Chen et al. [17] used the ARIMA [2] model to model the network traffic, extracted the multi-dimensional related features in the network traffic, and subtracted the real value from the predicted value of the multi-dimensional feature to construct the residual vector, and used OC-SVM to classify residual vector to realize anomaly detection of network traffic. Min et al. [18] first used the PCA algorithm to reduce the dimensionality of the time series, then used a sliding window to divide the time series and extract relevant features, finally used 1-SVM to detect anomalies in the sliced time series fragments.

In addition, decision tree algorithms are also widely used in anomaly detection. Zhou et al. [19] proposed an isolation forest algorithm for anomaly detection, which established multiple decision trees for multi-dimensional features to detect global outliers. Aryal et al. [20] improved the isolation forest algorithm to make it suitable for local anomaly detection. Liu [21] used the idea of random forest to extract hundreds of features from the labeled KPI data set, and trained the classifier through integrated learning. In response to the problem of data labeling, Zhao et al. [22] proposed a KPI sequence labeling framework Label-Less. Firstly, all candidate abnormal subsequences in KPI were screened by using isolated forest algorithm and setting an abnormal threshold, and then the similarity between all candidate sequences and manually selected abnormal sequences was calculated by using similarity alignment algorithm dynamic time warping (DTW), The candidate sequences with the highest similarity are marked as exceptions. The time of manual annotation can be reduced by 90%.

4.3 Method-Based Prediction

The method based on prediction mainly obtains the deviation degree by making the difference between the real value and the predicted value, and determines whether the data point is abnormal by the size of the deviation degree. The common prediction algorithms in time series include differential integrated moving average autoregressive model (ARIMA), Holt-Winters method (Holt-Winters) and prophet proposed by Facebook [23].

The ARIMA model is mainly used to predict short time series, and is only suitable for stationary series. However, the real sequence is generally a non-stationary sequence, so this model has certain limitations. The Holt-Winters algorithm is suitable for non-stationary series with linear trends and periodic fluctuations. The exponential smoothing method is used to fit the time series and make predictions. Similar to ARIMA, Holt-Winters can only predict short-term time series. The prophetic algorithm proposed by Facebook can automatically process outliers and missing values, and decompose the time series into trend, seasonal and holiday components, and fit the above components separately to predict the future trend of the time series.

To a certain extent, machine learning algorithms can make up for the shortcomings of heuristic rule-based methods in usability and universality. However, machine learning algorithms need to manually extract time series features, and the accuracy of anomaly detection directly depends on feature engineering. The high dimensionality of multi-dimensional time series brings greater challenges to extracting and constructing representative sequence features. In recent years, the academic community has proposed to apply the idea of deep learning to time series-oriented anomaly detection tasks, using models to learn the internal correlations of massive data, and automatically construct features to solve the limitations and limitations of the above-mentioned traditional methods insufficient.

5 Anomaly Detection Algorithm Based on Deep Learning

Deep learning is an extension of the field of machine learning. By learning the sample rules and internal representations in the data set, it has solved many pattern recognition problems. It has been applied to search recommendation, data mining, natural language processing and other researches field. At the same time, due to the high dimensionality and large amount of data in time series data, traditional outlier detection algorithms are no longer suitable for large-scale time series data sets. Chalapathy et al. [24] proposed the concept of deep anomaly detection through the idea of deep learning, the discriminative features in time series are represented and learned, and the features are automatically selected by using the model, which saves the step of manual feature selection by domain experts. However, the distinction between normal points and abnormal points in a data set is often relatively vague in most fields and may change. This kind of unclear boundary also brings challenges to deep anomaly detection methods, which often need to be analyzed for specific business.

At present, according to the principle of the method, deep anomaly detection can be divided into regression-based methods and dimensionality reduction methods. The following will focus on these two methods (Fig. 5).

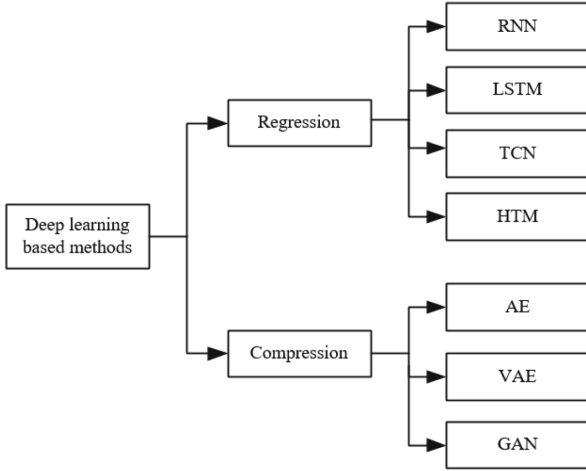


Fig. 5. Anomaly detection algorithms based on deep learning.

5.1 Method-Based Regression

One of the mainstream methods in the current time series anomaly detection task is to use the idea of regression and use a certain sequence prediction model to predict the value at $t + 1$ based on the observation value at the previous t time, and make the difference with the real value at that time to evaluate whether the time series at that moment is abnormal.

At present, several mainstream series prediction models based on deep learning mainly include recurrent neural network (RNN) [25], long-term and short-term memory artificial neural network (LSTM) [26], gated cyclic unit (Gru) [27], and time convolution network (TCN) proposed by Bai et al. [28] in 2018.

RNN is a type of recursive neural network that takes sequence data as input, recursively in accordance with the advancement direction of the sequence, and all cyclic unit nodes are connected in a chain. RNN can capture the temporal and spatial semantics in the time series, but it is easy to produce phenomena such as gradient disappearance during the training process, and the recursive training process cannot be parallelized, and the model convergence speed is slow. In order to solve problems such as the gradient disappearance of RNN, LSTM improves RNN by adding input gates, output gates, forget gates and memory units to the network, which can learn long-term dependencies in time series and record time series Important events with long intervals and delays. GRU is a variant of LSTM, which simplifies the network structure of LSTM, introduces update gates and reset gates, saves important features in the time series through the gate function, and ensures that the gradient is not lost during the training process. Compared with LSTM, GRU has reduced the number of parameters, which can accelerate model convergence. TCN is a newly proposed time series prediction model based on Convolutional Neural Network (CNN) [29] in recent years. It uses causal convolution to capture short-term sequence semantics, expanded convolution to capture long-term dependent

semantics, and finally passes through a layer of residual the difference network solves the problem of the disappearance of the gradient, and predicts the time series through the above ideas. As a variant of CNN, TCN is different from models such as RNN in that it can support parallel computing, so it can accelerate model training. The following will introduce in detail the related work applied to the above models in anomaly detection tasks.

Among them, in the RNN-based model, Thi et al. [30] and Bontemps et al. [31] regard the network intrusion detection task as a binary classification problem, and use RNN to model the sum of deviations of an entire time series to detect abnormal patterns in the data set. Banjanovic mehmedovic et al. [32] constructed a data-driven model based on neural network for real-time monitoring of thermal power plant system, and used MLP [33], RNN and probability and statistics models for comparison. Saurav et al. [34] analyzed the shortcomings of modeling historical data in offline environments for anomaly detection tasks. Due to the dynamic changes of real-time data in the real environment, the normal mode in the time series may change, which greatly reduces the accuracy of model detection. This paper improves RNN with the idea of incremental learning, integrates new data in the real production environment, detects abnormal points and change points in the time series according to the difference between the predicted value and the real value, and updates the RNN network parameters at the same time, So that the model can monitor the anomalies in the online environment in real time. Guo et al. [35] proposed an adaptive gradient learning method based on RNN for time series prediction tasks, which modeled the local features in the time series, and automatically weighted the loss gradient of new observations generated in real time to the existing historical data, so as to achieve the purpose of adaptive learning. Experiments are carried out on artificial data sets and real data sets, and the effect of the model is evaluated. Qin et al. [36] proposed an RNN autoencoder based on the attention mechanism, which can more accurately predict the long-term dependence in the time series.

As a variant of RNN, LSTM also has a very wide range of applications in time series anomaly detection tasks. Malhotra et al. [37] used the normal points in the data set to train the LSTM, and modeled the error between the predicted value and the true value of multiple points in a period of time as a multivariate Gaussian distribution, which was used to evaluate the possibility of abnormality at each time point. Sucheta et al. [38] applied similar ideas to the task of ECG signal detection. Donghyun et al. [39] introduced the concept of edge computing to the anomaly detection model based on LSTM, which can accelerate calculations and reduce network resource consumption. The proposed system LiReD has been applied to real-time monitoring of industrial environments and achieved good performance. Hundman et al. [40] applied LSTM to the spacecraft anomaly monitoring task. In this paper, the time series data generated by each sensor are modeled separately, and an unsupervised and parameterless anomaly threshold calculation method is proposed to set the anomaly limits. LSTM is also widely used in automobile control network [41], industrial Internet of things monitoring [37], network traffic monitoring [42] and other fields.

In addition, Fu et al. [43] used LSTM and GRU to predict traffic flow, and proved that deep learning can achieve better results than ARIMA and other traditional statistical models through experiments. Mohsin et al. [44] proposed an anomaly detection framework DeepAnT, the DeepAnT is divided into two parts, namely the prediction component and the anomaly detection component. The prediction component refers to the idea of the TCN convolutional network, the output prediction value is input to the detection component, and the Euclidean distance [45] between the prediction value and the true value is used to determine whether an abnormality occurs at this moment. In the past two years, Cui et al. [46] proposed a new sequence prediction model-Hierarchical Time Memory Network (HTM), which is based on a bionic design and was subsequently used in time series anomaly detection tasks [47–49].

5.2 Method-Based Dimension Reduction

When a system is jointly monitored by multiple sensors, a large number of monitoring KPIs will be generated during the monitoring process. These KPIs not only have a very long time dimension, but may also influence each other internally, and have very complex correlation characteristics. These factors bring difficulties to the process of data mining, which jointly restricts the accuracy of anomaly detection algorithm.

In view of the above problems, it is easy to think of using the idea of dimension reduction to solve the problem of high data dimension. Among them, Principal Component Analysis (PCA) [50], a typical algorithm of dimensionality reduction, extracts the linear uncorrelated components of a set of variables through orthogonal transformation to achieve the purpose of dimensionality reduction. Based on this idea, deep learning can be used to learn the dimensionality reduction representation method of the normal pattern in the time series, and the dimensionality reduction feature vector can be reconstructed to restore to the original dimension, which is defined according to the reconstruction error of the input and output sequences. Whether the sequence is abnormal. Since data labels are not required, the method based on dimensionality reduction is actually an unsupervised method. A prerequisite of this method is that there are structural differences between the normal sequence and the abnormal sequence, that is, the normal sequence can be restored by the model, and the abnormal sequence will produce larger reconstruction error. Anomaly detection algorithms that use the idea of reconstruction error mainly include Autoencoder (AE) [51] and its variant Variational Autoencoder (VAE) [52] and Generative Adversarial Network (GAN) [53].

The autoencoder uses the input information as the learning target to perform characterization learning on the input information [54]. In terms of structure, the autoencoder is divided into two parts: an encoder and a decoder. The encoder encodes the input, and the output dimension is generally much smaller than the input dimension; The decoder decodes it and restores it to the same dimension as the input. VAE is a variant of autoencoder, which is a generation model like GAN. The goal is to build a model that generates target data X from latent variable Z and learn the transformation between distributions. GAN is divided into generator and discriminator in structure, and learns the feature representation through mutual games and joint training between the two. The related work applied to the above model will be described in detail below.

Sakurada et al. [55] used the idea of dimensionality reduction to apply the autoencoder to the field of anomaly detection for the first time, and compared it with traditional dimensionality reduction methods such as PCA and kernel-PCA through experiments. The experiments proved that the autoencoder can improve the accuracy of the anomaly detection model. Kieu et al. [56] divided the time series into multiple sliding windows, extracted eight-dimensional features for each time window, and spliced them with external semantic information, and the reconstruction error is trained to the minimum by inputting to LSTM-AE and CNN-AE. Meng et al. [57] expanded on the work of [56] and combined time convolutional networks with autoencoders to detect abnormal points in the time series of the Cyber Physical Social System (CPSS). Zhang et al. [58] proposed a multi-angle convolutional recursive autoencoder (MSCRED), which first calculates the feature matrix for the multi-dimensional features of each moment in the time dimension, and then uses CNN-AE and ConvLSTM to learn the spatial semantics of the time series. Features and temporal semantic features, the model can locate anomalies based on anomalous point detection, and classify anomalies. Kieu et al. [59] used the idea of ensemble learning and proposed an autoencoder based on sparse RNN, which trains multiple AE models by changing the RNN network structure, and finally uses the median of the reconstruction error of each model output as the classification result. This method can solve the over-fitting problem in the deep neural network training process. Luo et al. [60] introduced the concept of cloud computing on the basis of AE, which can efficiently detect anomalies in wireless sensor networks in a distributed environment. At the same time, the anomaly detection algorithms based on autoencoders also have been applied in the fields of energy consumption monitoring [61–63], aircraft monitoring [64], and network intrusion detection [65].

For the generative model, Kim et al. [66] used CNN-VAE to detect the timing anomalies of edge devices in industry IOT big data environment. Guo et al. [67] proposed a GRU-based Gaussian Mixture Variational Autoencoder (GGM-VAE), by learning the temporal and spatial semantic features in multi-dimensional time series, and setting a reconstruction error threshold to define whether an abnormality occurs at that moment. Park et al. [68] used similar ideas to apply LSTM-VAE to robots in behavioral anomaly detection. Xu et al. [69] proposed a VAE-based anomaly detection framework DONUT, which uses evidence lower bounds, missing value injection, and Markov Chain Monte Carlo method MCMC to improve model detection accuracy, it is mainly used in Internet company's Abnormal detection of monitoring indicators KPI.

Similar to VAE, GAN is also a generative model. Zenati et al. [70] applied GAN to the field of sequence anomaly detection for the first time, and evaluated the effect of the model on images and network intrusion detection data sets. Li et al. [71] considered the potential interaction of time series data generated by multi-sensor in industrial environment, used LSTM-RNN as the generator in GAN to learn the common distribution of multi-dimensional time series, and detected outliers according to the results of discriminator and reconstruction error of generator. Lim et al. [72] aimed at the imbalance problem of anomaly detection data set, and used GAN to generate artificial samples to expand the data set and improve the detection effect of anomaly algorithms.

Relying on the advantages of strong learning ability, wide coverage and strong adaptability, the deep learning method automatically learns the intrinsic correlation and essence of massive data through the model, and automatically constructs representative excellent features as the decision basis of the classifier. It avoids the time-consuming and labor-consuming human feature engineering link, and further improves the accuracy of the algorithm, and effectively makes up for the shortcomings of traditional methods, and has been applied to multi-dimensional time series oriented anomaly detection task. However, the current anomaly detection algorithms based on deep learning are still immature, and many algorithms and models are still in the offline detection stage. When facing the actual production environment, there are still shortcomings such as high delay of anomaly detection methods and poor model adaptation.

6 Summary

This paper summarizes the anomaly detection methods of multi-dimensional time series, and the anomaly detection algorithms are summarized in Table 2. With the rapid development of science and technology, the complexity of industrial system architecture is also increasing. At the same time, various system monitoring indicators can be collected in real time through software, sensors and other media, thus forming a large-scale multi-dimensional time series data. Through these monitoring indicators, the real-time operating status of the system can be analyzed and evaluated, and real-time response can be achieved when abnormalities are found, and economic losses can be reduced as much as possible.

In the task of time series anomaly detection, the method of setting thresholds and box plot statistics is simple and intuitive, and can achieve better results on small sample data. When the data sample is further expanded and the time series dimension rises, the method cannot capture the spatiotemporal semantics in the sequence, which leads to the relatively high false alarm rate and missing alarm rate of traditional methods. Deep learning methods use data normalization and sliding windows to perform data normalization and sliding windows on the original time series by learning the internal connections and laws of the data, constructs regression model and classification model to predict the data value of the future time series, and captures the time semantics in the series. At the same time, learn the feature representation of the sequence through the autoencoder network, and capture the spatial semantics in the sequence, which greatly improves the accuracy of the algorithm.

However, this also puts forward higher requirements for anomaly detection algorithms, the existing algorithms still have some shortcomings that need to be improved.

The Algorithm Training Time is Long. In most fields, especially in the industrial IOT environment, the data dimension of time series is very high. At the same time, the number of parameters in the deep learning model is also very large, resulting in a large amount of network resource consumption in the training process and a long training time of the model.

Table 2. Summary of anomaly detection algorithms for multi-dimensional time series.

Type	Idea	Reference	Algorithm
Rule based	Thresholding	Ref [6]	Box Plot
Machine learning based	Clustering	Ref [8–12]	KNN
		Ref [10–12]	DBSCAN
	Classification	Ref [16–18]	OC-SVM,MC-SVM
		Ref [19–22]	Isolation Forest
	Prediction	Ref [2, 17]	ARIMA
		Ref [23]	FB-Prophet
Ref [14]		Holt-Winters	
Deep learning based	Regression	Ref [25, 30–36]	RNN, S-RNN,Attention-RNN
		Ref [28, 44]	TCN
		Ref [26, 37–42]	LSTM,Attention-LSTM,GRU-LSTM
		Ref [46–49]	HTM
	Compression	Ref [51, 55–65]	AE, CNN-AE, LSTM-AE, RNN-AE
		Ref [52, 66–69]	VAE, CNN-VAE, GGM-VAE, LSTM-VAE
Ref [53, 70–72]		GAN	

The Adaptability of Model is Poor. The data volume of the monitoring indicators increases with time, the abnormal patterns in the time series may change with the upgrade of the system architecture and the number of server clusters. If the algorithm model is only trained based on historical data, the model may no longer be applicable and the detection accuracy rate will be greatly reduced when the above situation occurs. In view of the above problems, the concept of incremental learning should be introduced, however, there is relatively little work based on incremental learning in the field of anomaly detection.

The Universality of Algorithm is Low. At present, the anomaly detection algorithms proposed by existing work often perform well in specific scenarios or for a certain data set. There is no algorithm or model that can be applied to multiple fields. The algorithms in each field cannot be universal, and the universality and scalability of the model are poor.

Generally speaking, the rule-based method has been more mature and used in industrial production environments. However, their false positive rate and false negative rate are relatively high, which brings a lot of human workload to operation and maintenance personnel. Machine learning algorithm needs to construct features manually, and the data collection and annotation is time-consuming and labor-consuming, all of these factors will cause certain influence and deviation to the anomaly detection results. At the same time, the newly proposed methods based on deep learning in recent years can be

used in Improve the accuracy of the algorithm to a certain extent, but there are still difficulties in the actual implementation of the algorithm. Therefore, the future research on anomaly detection algorithms should be combined with the actual industrial production environment, and the real-time data collected in the production environment should be used as the standard to test the feasibility of the algorithm, so as to improve the practical application value of the model algorithm.

Acknowledgment. The research work is supported by National Key R&D Program of China (No. 2018YFB0804204).

References

1. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. *ACM Comput. Surv.* **41**(3), 1–58 (2009)
2. Box, G.E.P., Jenkins, G.M., Reinsel, G.C.: Time series analysis forecasting and control. *J. Time* **31**(2), 238–242 (1976). Rev. edn.
3. Romera, E., Bergasa, L.M., Arroyo, R.: Need data for driver behaviour analysis? Presenting the public UAH-DriveSet. In: *ITSC 2016*, pp. 387–392 (2016)
4. Aggarwal, C.C.: An introduction to outlier analysis. In: Aggarwal, C.C. (ed.) *Outlier Analysis*, pp. 1–34. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-47578-3_1
5. Grubbs, F.E.: Procedure for detecting outlying observations in samples. *Technometrics* **11**(1), 53 (1974)
6. Laurikkala, J., Juhola, M., Kentala, E., Lavrac, N., Miksch, S., Kavsek, B.: Informal identification of outliers in medical data. In: *Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology*, vol. 1, pp. 20–24 (2000)
7. Laptsev, N., Amizadeh, S., Flint, I.: Generic and scalable framework for automated time-series anomaly detection (2015)
8. Ramaswamy, S., Rastogi, R., Shim, K.: Efficient algorithms for mining outliers from large data sets. *ACM SIGMOD Rec.* **29**(2), 427–438 (2000)
9. Hautamaki, V., Karkkainen, I., Franti, P.: Outlier detection using k-nearest neighbour graph. In: *ICPR*, pp. 430–433 (2004)
10. Li, Z., Zhao, Y., Liu, R., et al.: Robust and rapid clustering of KPIs for large-scale anomaly detection. In: *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*. ACM (2018)
11. Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: *KDD*, vol. 96, no. 34, pp. 226–231 (1996)
12. Bu, J., Liu, Y., Zhang, S., et al.: Rapid deployment of anomaly detection models for large number of emerging KPI streams. In: *2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC)*. IEEE (2018)
13. Mahimkar, A., et al.: Rapid detection of maintenance induced changes in service performance. In: *Proceedings of the Seventh Conference on Emerging Networking EXperiments and Technologies*, ser. CoNEXT 2011, pp. 13:1–13:12. ACM, New York (2011). <http://doi.acm.org/10.1145/2079296.2079309>
14. Barford, P., Kline, J., Plonka, D., Ron, A.: A signal analysis of network traffic anomalies. In: *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement*, pp. 71–82. ACM (2002)

15. Yan, H., et al.: Argus: end-to-end service anomaly detection and localization from an ISP's point of view. In: 2012 Proceedings IEEE INFOCOM, pp. 2756–2760. IEEE (2012)
16. Manevitz, L.M., Yousef, M.: One-Class SVMs for document classification. *JMLR* **2**, 139–154 (2001)
17. Chen, X., Jiang, T., et al.: Network anomaly detector based on multiple time series analysis. *J. Sichuan Univ. (Eng. Sci. Edn.)* **49**(001), 144–150 (2017)
18. Min, H., Zhiwei, J., Ke, Y., et al.: Detecting anomalies in time series data via a meta-feature based approach. *IEEE Access* **1** (2018)
19. Liu, F.T., Ting, K.M., Zhou, Z.-H.: Isolation forest. In: *ICDM*. IEEE (2008)
20. Aryal, S., Ting, K.M., Wells, J.R., et al.: Improving iForest with relative mass (2014)
21. Liu, D., Zhao, Y., Xu, H., et al.: Opprentice: towards practical and automatic anomaly detection through machine learning. In: *Proceedings of the 2015 Internet Measurement Conference*, pp. 211–224. ACM Press, New York (2015)
22. Zhao, N., Zhu, J., Liu, R., et al.: Label-less: a semi-automatic labelling tool for KPI anomalies. In: *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*. IEEE (2019)
23. Taylor, S.J., Letham, B.: Forecasting at scale. *Am. Stat.* (2017)
24. Chalapathy, R., Chawla, S.: Deep learning for anomaly detection: a survey (2019)
25. Williams, R.J.: Complexity of exact gradient computation algorithms for recurrent neural networks. Technical report, Technical report NU-CCS-89-27. Northeastern, Boston (1989)
26. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
27. Cho, K., et al.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014)
28. Bai, S., Kolter, J.Z., Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling (2018)
29. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012)
30. Nguyen Thi, N., Cao, V.L., Le-Khac, N.A.: One-class collective anomaly detection based on LSTM-RNNs. In: Hameurlain, A., Küng, J., Wagner, R., Dang, T., Thoai, N. (eds.) *Transactions on Large-Scale Data- and Knowledge-Centered Systems XXXVI*. LNCS, vol. 10720, pp. 73–85. Springer, Berlin, Heidelberg (2017). https://doi.org/10.1007/978-3-662-56266-6_4
31. Bontemps, L., Cao, V.L., McDermott, J., Le-Khac, N.A.: Collective anomaly detection based on long short-term memory recurrent neural networks. In: Dang, T., Wagner, R., Küng, J., Thoai, N., Takizawa, M., Neuhold, E. (eds.) *FDSE 2016*. LNCS, vol. 10018, pp. 141–152. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48057-2_9
32. Banjanovic-Mehmedovic, L., Hajdarevic, A., Kantardzic, M., Mehmedovic, F., Džananovic, I.: Neural network-based data-driven modelling of anomaly detection in thermal power plant. *Automatika* **58**(1), 69–79 (2017)
33. Pal, S.K., Mitra, S.: Multilayer perceptron, fuzzy sets, classification (1992)
34. Saurav, S., et al.: Online anomaly detection with concept drift adaptation using recurrent neural networks. In: *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, pp. 78–87. ACM (2018)
35. Guo, T., Xu, Z., Yao, X., Chen, H., Aberer, K., Funaya, K.: Robust online time series prediction with recurrent neural networks. In: 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 816–825. IEEE (2016)
36. Qin, Y., Song, D., Chen, H., et al.: A dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint arXiv:1704.02971* (2017)
37. Zhang, W., et al.: LSTM-based analysis of industrial IoT equipment. *IEEE Access* **6**, 23551–23560 (2018)

38. Chauhan, S., Vig, L.: Anomaly detection in ECG time signals via deep long short-term memory networks. In: IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 1–7. IEEE (2015). 36678
39. Park, D., Kim, S., An, Y., Jung, J.-Y.: LiReD: a light-weight real-time fault detection system for edge computing using LSTM recurrent neural networks. *Sensors* **18**(7), 2110 (2018)
40. Hundman, K., Constantinou, V., Laporte, C., et al.: Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 387–395 (2018)
41. Taylor, A., Leblanc, S., Japkowicz, N.: Anomaly detection in automobile control network data with long short-term memory networks. In: 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 130–139. IEEE (2016)
42. Cheng, M., et al.: MS-LSTM: A multi-scale LSTM model for BGP anomaly detection. In: 2016 IEEE 24th International Conference on Network Protocols (ICNP), pp. 1–6. IEEE (2016)
43. Fu, R., Zhang, Z., Li, L.: Using LSTM and GRU neural network methods for traffic flow prediction. In: Youth Academic Annual Conference of Chinese Association of Automation (YAC), pp. 324–328. IEEE (2016)
44. Munir, M., Siddiqui, S.A., Dengel, A., Ahmed, S.: DeepAnT: a deep learning approach for unsupervised anomaly detection in time series. *IEEE Access* **7**, 1991–2005 (2018)
45. Danielsson, P.E.: Euclidean distance mapping. *Comput. Graph. Image Process.* **14**(3), 227–248 (1980)
46. Cui, Y., Ahmad, S., Hawkins, J.: Continuous online sequence learning with an unsupervised neural network model. *Neural Comput.* **28**(11), 2474–2504 (2016). <http://arxiv.org/abs/1512.05463>
47. Ahmad, S., Purdy, S.: Real-time anomaly detection for streaming analytics. [arXiv:1607.02480](https://arxiv.org/abs/1607.02480) [cs], July 2016
48. Ahmad, S., Lavin, A., Purdy, S., Agha, Z.: Unsupervised real-time anomaly detection for streaming data. *Neurocomputing* **262**, 134–147 (2017). <http://www.sciencedirect.com/science/article/pii/S0925231217309864>
49. Wu, J., Zeng, W., Yan, F.: Hierarchical temporal memory method for time-series-based anomaly detection. *Neurocomputing* **273**, 535–546 (2018). <http://www.sciencedirect.com/science/article/pii/S0925231217313887>
50. Baldi, P., Hornik, K.: Neural networks and principal component analysis: learning from examples without local minima. *Neural Netw.* **2**(1), 53–58 (1989)
51. Vincent, P., Larochelle, H., Bengio, Y., et al.: Extracting and composing robust features with denoising autoencoders. In: International Conference on Machine Learning. ACM (2008)
52. Kingma, D.P., Welling, M.: Auto-encoding variational bayes (2014)
53. Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
54. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(8), 1798–1828 (2013)
55. Sakurada, M., Yairi, T.: Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis, p. 4. ACM (2014)
56. Kieu, T., Yang, B., Jensen, C.S.: Outlier detection for multidimensional time series using deep neural networks. In: 2018 19th IEEE International Conference on Mobile Data Management (MDM). IEEE (2018)
57. Meng, C., Jiang, X.S., Wei, X.M., et al.: A Time convolutional network based outlier detection for multidimensional time series in cyber-physical-social systems. *IEEE Access* **8**, 74933–74942 (2020)
58. Zhang, C., Song, D., Chen, Y., et al.: A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data (2018)

59. Kieu, T., Yang, B., Guo, C., et al.: Outlier detection for time series with recurrent autoencoder ensembles. In: Twenty-Eighth International Joint Conference on Artificial Intelligence IJCAI-19 (2019)
60. Luo, T., Nagarajany, S.G.: Distributed anomaly detection using autoencoder neural networks in WSN for IoT. In: 2018 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE (2018)
61. Yuan, Y., Jia, K.: A distributed anomaly detection method of operation energy consumption using smart meter data. In: 2015 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), pp. 310–313. IEEE (2015)
62. Araya, D.B., Grolinger, K., ElYamany, H.F., Capretz, M.A.M., Bitsuamlak, G.: An ensemble learning framework for anomaly detection in building energy consumption. *Energy Build.* **144**, 191–206 (2017)
63. Fan, C., Xiao, F., Zhao, Y., Wang, J.: Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data. *Appl. Energy* **211**, 1123–1135 (2018). <http://www.sciencedirect.com/science/article/pii/S0306261917317166>
64. Fu, X., Luo, H., Zhong, S., Lin, L.: Aircraft engine fault detection based on grouped convolutional denoising autoencoders. *Chin. J. Aeronaut.* **32**, 296–307 (2019)
65. Filonov, P., Lavrentyev, A., Vorontsov, A.: Multivariate industrial time series with cyber-attack simulation: fault detection using an LSTM-based predictive data model. arXiv preprint [arXiv:1612.06676](https://arxiv.org/abs/1612.06676) (2016)
66. Kim, D., et al.: Squeezed convolutional variational AutoEncoder for unsupervised anomaly detection in edge device industrial internet of things. In: 2018 International Conference on Information and Computer Technologies (ICICT), pp. 67–71, December 2018
67. Guo, Y., Liao, W., Wang, Q., Yu, L., Ji, T., Li, P.: Multidimensional time series anomaly detection: a GRU-based Gaussian mixture variational autoencoder approach. In: Asian Conference on Machine Learning, pp. 97–112 (2018)
68. Park, D., Hoshi, Y., Kemp, C.C.: A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder. *IEEE Robot. Autom. Lett.* **3**(3), 1544–1551 (2018)
69. Xu, H., Chen, W., Zhao, N., et al.: Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications. In: Proceedings of the 2018 World Wide Web Conference, pp. 187–196 (2018)
70. Li, D., Chen, D., Goh, J., Ng, S.: Anomaly detection with generative adversarial networks for multivariate time series. arXiv preprint [arXiv:1809.04758](https://arxiv.org/abs/1809.04758) (2018)
71. Li, D., Chen, D., Jin, B., Shi, L., Goh, J., Ng, S.-K.: MAD-GAN: multivariate anomaly detection for time series data with generative adversarial networks. In: Tetko, I.V., Kůrková, V., Karpov, P., Theis, F. (eds.) ICANN 2019. LNCS, vol. 11730, pp. 703–716. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30490-4_56
72. Lim, S.K., Loo, Y., Tran, N.T., et al.: Doping: Generative data augmentation for unsupervised anomaly detection with GAN. In: 2018 IEEE International Conference on Data Mining (ICDM), pp. 1122–1127. IEEE (2018)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

