Research on Time Series Anomaly Detection: Based on Deep Learning Methods

Jiaqi Zhou
Automation college, Nanjing University of Posts and Telecommunications, Nanjing, 21000, China
dreamer_zjq@163.com

Abstract: Time series anomaly detection has always been an important research direction. The early time series anomaly detection methods are mainly statistical methods and machine learning methods. With the powerful functions of deep neural network being continuously mined by researchers, the effect of deep neural network in anomaly detection task has been significantly better than the traditional methods. In view of the continuous development and application of deep neural networks such as transformer and graph neural network (GNN) in time series anomaly detection in recent years, the body of research lacks a comparative evaluation of deep learning methods. The evaluation is conducted on public datasets. By analyzing the evaluation criteria, this paper discusses the performance of each model, as well as the problems and development direction in the field of time series anomaly detection in the future. This study found that in the time series anomaly detection task, transformer is suitable for dealing with long-time series prediction, and studying the graph structure of time series may be the best way to deal with time series anomaly detection in the future.

1. Introduction
Time series is a series formed by sorting the values of some statistical indicators in chronological order. There are two major problems in time series analysis, time series prediction and time series anomaly detection. The prediction of time series is to predict the possible level in the next period of time or several years by analogy or extension according to the development process, direction and trend reflected in the time series. Anomaly detection of time series is to check whether the current data deviates significantly from the normal situation through historical data analysis. This paper mainly focuses on the methods and applications of time series anomaly detection but time series prediction can also be used as a means of time series anomaly detection.

Time series anomaly detection is important in industrial big data analysis, medical and health big data analysis, financial big data analysis, and even in the field of aerospace. With the development of information technology, the scale of time series data increases exponentially, and the difficulty of time series anomaly detection is increasing. In the past decade, Deep learning has achieved great success and is widely used in the field of anomaly detection. Many researchers try to use various methods to apply deep neural network to time series anomaly detection, which greatly promotes the development of time series anomaly detection.

Therefore, this paper mainly presents various application methods of deep neural network in time series anomaly detection. This includes not only mature neural networks such as Convolution Neural
Networks (CNNs) and Long-Short Term Memory (LSTMs), but also some neural networks that need more exploration, such as transformer and GNN. This paper selects a series of representative methods and compares them as much as possible. The contribution of this work is to propose some possible improvement directions for the shortcomings of the method and summarize the advantages of the method, which can be used as a reference for follow-up research.

The rest of the paper is organized as follows: in Section 2, the author reviews three different time series anomaly detection methods and introduces them according to the types of models. In Section 3, the author compares and evaluates each model with evaluation indexes, and puts forward some open problems for future research.

Anomaly detection approaches in time series

In this paper, the author only discuss the deep learning methods of time series anomaly detection, which are mainly divided into the following three categories. Supervised learning method tells the model what normal data points and abnormal data points look like by labeling data, and then trains the classification model, also known as discriminative algorithm. Unsupervised learning method, training and modeling for normal data, and then identifying abnormal points through high reconstruction error, also known as generative algorithm. The probability distribution of data is modeled, and then abnormal points are identified according to the correlation between sample points and very low probability, such as Deep Autoencoding Gaussian Mixture Model (DAGMM) [1].

2. Models

2.1 Supervised learning method

Time series is an image, which shows that an ordered sequence changes with time. Supervised learning tells the model what is abnormal time series and what is normal time series by labeling data. There are several common time series anomalies. Innovative outlier (IO), which causes the interference of outliers, not only acts on time T, but also affects all observations of the sequence after time T. Additive outlier (AO), which causes the interference of this outlier, only affects the sequence value at the time T at which the interference occurs, but does not affect the sequence value after that time. Level shift (LS), the interference caused by this outlier is that at a certain time T, the structure of the system has changed and continues to affect all behaviors after time T. In the sequence, it often shows that the sequence mean before and after time t has a horizontal displacement. Temporary change (TC): the interference causing this outlier has a certain initial effect when the interference occurs at time T, and then decays exponentially with time according to the attenuation factor.

Some obvious exceptions are not listed, such as missing and minor. Generally, the anomaly detection algorithm should mark each time point as abnormal or normal, or predict the signal of a point, and measure whether the difference between the real value and the predicted value is large enough to treat it as abnormal. In the selection of network, CNN has obvious advantages in the task of image classification. Tailai Wen et al [2] proposed a time series segmentation method based on CNN for anomaly detection. In addition, a transfer learning framework is proposed, which draws the model in advance on the large-scale synthetic univariate time series data set, and then fine tunes it with the small-scale, univariate or multivariable data set of previously invisible exception categories.

2.2 Unsupervised learning method

There are many researches on the application of unsupervised learning in time series anomaly detection. This paper focuses on the application of LSTM, transformer, GNN model and some practical models.

2.2.1 LSTM

Looking back on unsupervised learning method in recent years, the neural network that may be most suitable for processing series data is LSTM. If properly constructed, this recurrent neural network can model and realize the most complex dependencies in time series, and this method is also effective if there are multiple time series coupled with each other. Tung Kieu et al [3] proposed an autoencoder (AE)
framework based on 2D convolutional neural network and LSTM model. Firstly, a method is proposed to generate statistical features to enrich the feature space of the original time series, and then the autoencoder is used to reduce the dimension of the extended time series features. By comparing the time series reconstructed by AE with the extended time series, we can detect whether there are outliers in the time series. The core idea is that compared with normal data, abnormal points are more difficult to be represented by low-dimensional feature space. Therefore, when abnormal points are encountered, the reconstruction error of AE will become large. Therefore, the reconstruction error can be used to judge whether there are outliers in the data. Figure 1 presents the AE network structure enhanced based on the statistical characteristics of time series. Daehyung Park et al [4] proposed a detector based on LSTM-based Variational Autoencoder (LSTM-VAE). This detector can also be used in time series anomaly detection. For a batch of data, the goal of generating the model is to learn to obtain a distribution \( P(x) \), so that the distribution is very close to the real distribution \( P_{gt}(x) \) of the data. However, we cannot directly obtain \( P(x) \). Since it is in the form of probability diagram, we can use the following model to generate \( x \) by introducing a hidden layer variable \( z \), \( P(x) \) can be written as:

\[
p(x) = \sum_z p(x|z)p(z)
\]

Finally, VAE also realizes the function similar to AE. Input a sequence to get an implicit variable, and then reconstruct the implicit variable into the original input. The difference is that VAE learns the distribution of hidden variables (allowing hidden variables to have certain noise and randomness), so it can have the effect of similar regularization to prevent over fitting.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{ae_network_structure.png}
\caption{AE network structure (2DCNN, LSTM) based on statistical property enhancement of time series}
\end{figure}

\subsection{Transformer}
Transformer is usually used for time series prediction, Haoyi Zhou et al [5] proposed to use transformer to realize long sequence time-series forecasting. Transformer has the ability of long sequence time-series forecasting, but it needs to be optimized in terms of computing time complexity and spatial complexity and how to strengthen the correlation between long-time series input and output. Three deficiencies are mentioned in the paper. Square order computational time complexity of self-attention mechanism, transformer usually stacks multi-layer networks, resulting in memory bottleneck, and step by step decoding prediction makes the reasoning speed slow.
In their paper, they proposed ProbSparse self-attention, self-attention distilling and generative style decoder to solve the above problems respectively. Therefore, transformer can be more effectively used in the prediction of long-time series. For the time series anomaly detection task, many researches need transformer to do long-time series prediction, such as GTA proposed by Zekai Chen et al [6].

2.2.3 GNN
Firstly, for multivariable time series, we can regard it as a matrix $X \in \mathbb{R}^{k \times n}$, which is composed of $k$ variables and $n$ moments. Most anomaly detection methods use normal data for modeling, and put the test data into the trained model to calculate the prediction error or reconstruction error. If the given threshold is exceeded, it is considered that an exception has occurred. Zhao et al [7] considered the combination of GNN from two dimensions of data matrix. One is that from the perspective of variables, a time series (single variable) can correspond to a node on the graph. The other is that from the perspective of time, the data vector (multivariable) at the same time can correspond to a node on the graph.

GDN [8] mined the connection relationship between variables in multivariable time series, that is, each time series is regarded as a node on the graph, but the connection relationship between nodes is learned, rather than simply assuming that it is a fully connected graph. Both papers use graph attention network for feature extraction.

2.2.4 Other models
Deep SVDD [9] used a neural network to map the samples from the input space to the output space, so that the normal samples gather near the center of the class in the hidden space. During the test, if a sample is too far from the center of the normal class, it is considered as abnormal. Ideally, normal samples and abnormal samples are clustered into two clusters in the output space. MAD-GAN [10] used generative adversarial networks (GAN) in time series anomaly detection, takes LSTM as the basic model to capture dependencies among different time series, and embeds it into the framework of GAN. USAD [11] proposed a codec for adversarial training framework, which combines the advantages of AE and adversarial training and makes up for its shortcomings.

2.2.5 DAGMM
DAGMM [1] is also an unsupervised learning method, but its processing is special. Figure 2 is a diagram given in paper.

![Figure 2. Low-dimensional representations for samples from a private cybersecurity dataset](image)

In Fig. 2, the horizontal axis is the coordinate of compressing data into one-dimensional space by using the autoencoder, and the vertical axis is the reconstruction error of data points. In the figure, blue
points are normal data points and red points are abnormal data points. Obviously, we can see from the figure that the abnormal data points in the upper right corner can be distinguished from other points through reconstruction error, while the abnormal data points in the lower left corner can be clearly distinguished from other points in the low dimensional feature space. DAGMM connects GMM to AE model, and finally uses the probability value output by GMM to judge whether the sample is abnormal.

3. Findings
In this chapter, supervised model and unsupervised model will still be discussed separately.

3.1 Supervised model
Through the exploration of researchers, CNN has been outstanding in the supervised image classification task with high accuracy. However, there are several problems to be solved in the anomaly detection task of time series. On the one hand, it is a common problem of supervised learning, which requires a large amount of data and data annotation. On the other hand, the impact of outliers in time series may last for a short time, which means that this method can not detect outliers in long time series. Moreover, the accuracy of the classification depends greatly on the quality of the dataset. Therefore, the supervised method is suitable for short periodic time series in specific scenarios.

3.2 Unsupervised model

3.2.1 Model comparison
Integrating all the mentioned models, referring to the performance of the models in Secure Water Treatment (SWaT) [12], Water Distribution (WADI) [13] and Mars Science Laboratory Rover (MSL) [14] public datasets in their paper, the models are compared and evaluated by using precision (P), recall (R) and F1 score (F1) evaluation criteria in Table 1. The formula is as follows:

\[
\text{Precise} = \frac{TP}{TP + FP} \quad (2)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (3)
\]

\[
\text{F1 score} = 2 \times \frac{\text{Precise} \times \text{Recall}}{\text{Precise} + \text{Recall}} \quad (4)
\]

Among them, TP is true positive, indicating the number of real abnormalities detected, FP is false positive, indicating the number of detection errors, and FN is false negative, which is an undetected abnormality.

| Methods   | Model | SWaT P | R | F1 | WADI P | R | F1 | MSL P | R | F1 |
|-----------|-------|--------|---|----|--------|---|----|-------|---|----|
| AE        | AE    | 0.9913 | 0.7040 | 0.8233 | 0.3970 | 0.3220 | 0.8535 | 0.9748 | 0.8792 |
| LSTM-VAE  | VAE   | 0.7123 | 0.9258 | 0.8051 | 0.4632 | 0.3220 | 0.3799 | 0.8599 | 0.9756 | 0.8537 |
| GTA       | Transformer | 0.9483 | 0.8810 | 0.9134 | 0.8391 | 0.8361 | 0.9104 | 0.9117 | 0.9111 |
| GDN       | GNN   | 0.9585 | 0.9142 | 0.9358 | 0.8562 | 0.8541 | 0.8552 | 0.8292 | 0.9919 | 0.9033 |
| MAD-GAN   | GAN   | 0.9897 | 0.6374 | 0.7754 | 0.4144 | 0.3392 | 0.3601 | 0.8517 | 0.8991 | 0.8747 |
| USAD      | VAE   | 0.9870 | 0.7402 | 0.8460 | 0.6451 | 0.3220 | 0.4296 | 0.8810 | 0.9786 | 0.9109 |
| DAGMM     | AE    | 0.8292 | 0.7674 | 0.7971 | 0.2228 | 0.1976 | 0.2094 | 0.7562 | 0.9803 | 0.8112 |
SWAT is a low dimensional large dataset, WADI is a high dimensional large dataset, and MSL is a low dimensional small dataset. As can be seen from table 1, VAE’s comprehensive ability in time series anomaly detection tasks is similar to AE, while GNN and transformer have obvious advantages in high-dimensional and large amount of data time series tasks. GNN and transformer are also outstanding in terms of comprehensive capabilities. GTA not only uses transformer for time series prediction, but also uses the properties of graph structure. This means that graph structure has great research significance in anomaly detection of time series.

3.2.2 Performance summary of anomaly detection methods

(1) AE, DAGMM
The ordinary AE model is suitable for two classifiers, which is not affected by the imbalance of data categories. It is easy to train, fast convergence and good generalization performance, but it is slightly insufficient in accuracy. DAGMM can be used as multiple classifiers through continuous dimensionality reduction. However, due to the unbalanced classification of data sets, DAGMM has slow convergence, depends on tuning parameters and a priori probability, and has poor generalization performance. It may have a satisfactory effect on a single dataset, but it needs a lot of work to migrate to other datasets. Both of them are not suitable for high dimensional time series.

(2) LSTM-VAE, MAD-GAN, USAD
The implementation function of LSTM-VAE is similar to AE, but it has the function of similar regularization to prevent over fitting. MAD-GAN has many super parameters, it is difficult to adjust the parameters of the model, the generalization ability is poor, but the training speed is fast. USAD uses adversarial training while VAE based, so as to isolate anomalies while providing fast training. Similarly, they are not applicable to high-dimensional time series.

(3) GTA, GDN
Both of their models make use of the graph structure characteristics of time series. GTA learns the connection relationship of graph structure and uses transformer to do the time series prediction task. GDN is to mine the connection relationship between variables in multivariable time series, and treat each time series as a node on the graph, similarly learns the connection relationship between nodes, and directly uses graph attention network to extract features. These two methods not only perform well in low-dimensional data sets, but also have accuracy that can not be achieved in high-dimensional data sets compared with other models. However, GNN cannot handle dynamic graphs well. When a node needs to be added or deleted, that is, when the number of variables changes, GNN cannot adapt to this change.

3.2.3 Open problems
Although the researchers have made great achievements in time series anomaly detection, there are still some problems that have not been well solved. In this chapter, the author will mainly raise some questions about unsupervised models for future studies:
Real systems usually produce data continuously, and exceptions are not invariable. Therefore, manual intervention can be carried out to correct the test results, and then the correct labels provided by experts can be fed back to the model, so that the model can actively learn the changes of data and obtain more accurate results.
In the real case, time series analysis may need to be analyzed under different variables. The current model is difficult to deal with the time series with changing variables.
Time series has a wide range of application scenarios and many types of anomalies. In reality, the frequency of anomalies in most cases will be very small, and the prediction of aperiodic signals by the model is difficult to achieve very accurate, which leads to some special anomalies that cannot be detected by prediction and other methods.

4. Conclusion
Time series anomaly detection is of great significance in industry, finance, aerospace and other fields. With the exponential growth of time series in dimension and quantity, it is unrealistic to judge the
anomaly of time series artificially, and every anomaly of time series may lead to serious consequences. Therefore, The accuracy of time series anomaly detection model needs to be further improved. The model with outstanding advantages in accuracy is mainly deep neural network, which is also the goal that researchers need to continue to explore.

In this paper, ten widely used models and model related anomaly detection methods are reviewed, and the performance of the model are evaluated. By analyzing the characteristics of the model and comparing each model on the public datasets, this paper puts forward the suitable application scenario of each model, and points out the advantages and disadvantages of each model in the application of time series anomaly detection. However, in the research process, due to the direct use of the numerical indicators provided in the paper, the data processing methods and training objectives of different research methods may be different, so the data may not be completely accurate, but it still reflects the performance of the model. Now, the amount of data in many industrial and commercial industries is increasing. Both data and labels are increasing exponentially, resulting in the increasing demand for unsupervised learning. Therefore, the author puts forward some improvement points and possible research directions for unsupervised learning model in the future.

Acknowledgments
Firstly, I would like to show my deepest gratitude to my teachers and professors in my university, who have provided me with valuable guidance in every stage of the writing of this thesis. Further, I would like to thank all my friends and parents for their encouragement and support. Without all their enlightening instruction and impressive kindness, I could not have completed my thesis.

Reference:
[1] Zong, B., Song, Q., Min, M. R., Cheng, W., Lumezanu, C., Cho, D., & Chen, H. (2018, February). Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In International conference on learning representations.
[2] Wen, T., & Keyes, R. (2019). Time series anomaly detection using convolutional neural networks and transfer learning. arXiv preprint arXiv:1905.13628.
[3] Kieu, T., Yang, B., & Jensen, C. S. (2018, June). Outlier detection for multidimensional time series using deep neural networks. In 2018 19th IEEE International Conference on Mobile Data Management (MDM) (pp. 125-134). IEEE.
[4] Park, D., Hoshi, Y., & Kemp, C. C. (2018). A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder. IEEE Robotics and Automation Letters, 3(3), 1544-1551.
[5] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021, May). Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of AAAI.
[6] Chen, Z., Chen, D., Zhang, X., Yuan, Z., & Cheng, X. (2021). Learning Graph Structures with Transformer for Multivariate Time Series Anomaly Detection in IoT. IEEE Internet of Things Journal.
[7] Zhao, H., Wang, Y., Duan, J., Huang, C., Cao, D., Tong, Y., ... & Zhang, Q. (2020, November). Multivariate time-series anomaly detection via graph attention network. In 2020 IEEE International Conference on Data Mining (ICDM) (pp. 841-850). IEEE.
[8] Deng, A., & Hooi, B. (2021, February). Graph neural network-based anomaly detection in multivariate time series. In Proceedings of the AAAI Conference on Artificial Intelligence(Vol. 35, No. 5, pp. 4027-4035).
[9] Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S. A., Binder, A., ... & Kloft, M. (2018, July). Deep one-class classification. In International conference on machine learning(pp. 4393-4402). PMLR.
[10] Li, D., Chen, D., Jin, B., Shi, L., Goh, J., & Ng, S. K. (2019, September). MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks.
In International Conference on Artificial Neural Networks (pp. 703-716). Springer, Cham.

[11] Audibert, J., Michiardi, P., Guyard, F., Marti, S., & Zuluaga, M. A. (2020, August). Usad: Unsupervised anomaly detection on multivariate time series. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 3395-3404).

[12] Mathur, A. P., & Tippenhauer, N. O. (2016). SWaT: a water treatment testbed for research and training on ICS security. 2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater). IEEE.

[13] Ahmed, C. M., Palleti, V. R., & Mathur, A. P. (2017, April). WADI: a water distribution testbed for research in the design of secure cyber physical systems. In Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks (pp. 25-28).

[14] Hundman, K., Constantinou, V., Laporte, C., Colwell, I., & Soderstrom, T. (2018, July). 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).