LSTM-GAN for Enhanced Anomaly Detection in Time Series Data

Talon Whitaker

Department of Computer Science, Texas A&M University, USA
talon.whitaker57@tamu.edu

Abstract: Anomaly detection (AD) is crucial for identifying abnormal patterns in time series data that deviate from expected behavior, potentially preventing serious failures. Traditional methods, such as Statistical Process Control (SPC), require extensive human knowledge for setting model assumptions. Unsupervised machine learning approaches have since evolved, employing clustering and predictive modeling to identify anomalies. This paper introduces the LSTM-GAN model, which leverages Long Short-Term Memory (LSTM) networks as the backbone of both the generator and discriminator, and utilizes Wasserstein distance for improved measurement accuracy. Tested on several real-world datasets, the proposed model demonstrates superior performance in detecting anomalies in time series data.

Keywords: Time series data; Anomaly detection; GAN.

1. Introduction

Exceptions can be defined as abnormal patterns that do not conform to expected behavior. [1] With the rapid growth of time series data, abnormal data may lead to serious failures. Anomaly detection (AD) refers to automatic identification of abnormal phenomena mixed with a large number of normal data.[2]Therefore, anomaly detection can provide an opportunity to take action and solve problems before potential problems cause disasters.

Traditionally, various statistical methods have been proposed to improve the threshold, such as SPC [3]. If the monitoring data calculated on an instance exceeds the control limit, it will be identified as an exception. Such methods require a lot of human knowledge to set the prior assumptions of the model [4].

Later, some unsupervised machine learning methods were proposed. A common method is to divide time series data into subsequences with a certain length, and use clustering method to find outliers. The other is to learn a model to predict or reconstruct time series data, and compare the actual value with the predicted value or the reconstructed value. A high prediction error or reconstruction error calculated indicates the existence of anomalies [5].

With the rapid development of artificial neural networks in various fields, more and more methods based on deep learning have emerged. GAN is widely used in many fields, such as image generation, data balance, image recognition and video processing. [6] GAN can learn the real data distribution and generate false data similar to the real data.

In this paper, a temporal data anomaly detection model called LSTM-GAN is proposed. The structure of the rest of this paper is as follows: In section 2, the relevant technologies used are introduced, in section 3, the model structure proposed in this paper is introduced, in section 4, the relevant experiments conducted with the model proposed in this paper are introduced, and the last part is summarized.

2. Related technologies

The generative adversarial network is based on the scenario of game theory, [7] in which two participants compete with each other to achieve the goal of Nash equilibrium. The GAN network flow chart is as follows:
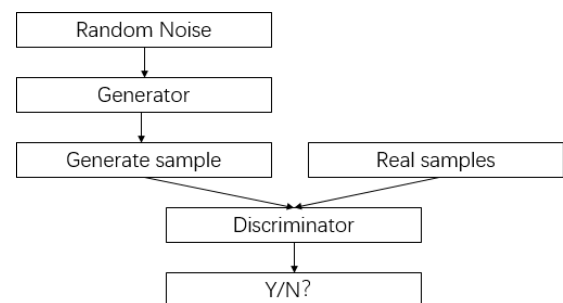

Figure 1. GAN network flow chart

GAN is composed of a generator G and a discriminator D. The random variable z from potential space Z is input into generator G to learn the distribution on real samples by mapping G(z). Generator G needs to generate generated samples that are as similar as possible to the distribution of real samples. The loss function of generator G is:

$$L_G = E_{z \sim p_{z(z)}}[\log(1 - D(G(z)))] \tag{1}$$

Input the real sample x or generated sample into discriminator D, and its training goal is to correctly distinguish between the real sample and generated sample.

After training, generator (G) has the ability to generate false data Discriminator (D) has the ability to distinguish between real data and false data.

Assuming $p_{data(x)}$ and $p_{data(x)}$ are known, the optimal discriminator function is:

$$D^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)} \qquad (2)$$

Substitute the optimal discriminator into formula 2, and the objective function becomes:

$$L(G|D^*) = 2D_{JS}(p_{data} \parallel p_g) - 2\log 2 \qquad (3)$$

When the discriminator is optimal, minimizing the JS difference between the actual distribution and the generated distribution is taken as the optimization goal of the generator. It will cause the JS divergence in the generator loss function to be equal to a constant when the real distribution and the generated distribution do not overlap or the overlapping part can be ignored, and the output of the discriminant network to all generated data is 0, the gradient disappears.

The training process of traditional generation countermeasure network is unstable and mode collapse may occur, which makes the training difficult and takes a long time.

In view of the problem of generator and discriminator confronting instability in the training process of GAN, Wasserstein distance was used instead of JS divergence to optimize training and generate countermeasures network. [8]Wasserstein distance is defined as follows:

$$W^{-1}(p_{data}, p_g) = \inf_{\delta \in \Gamma(p_{data}, p_g)} E_{(x,y) \sim \delta} \left[\parallel x - y \parallel \right] \qquad (4)$$

Since it is difficult to directly calculate the Wasserstein distance of two distributions, the Kantorovich Rubinstein dual theorem is adopted as the calculation form:

$$W^{-1}(p_{data}, p_g) = \frac{1}{K} \sup_{\parallel f \parallel_L \leq K} \left(E_{x \sim p_{data}}[f(x)] - E_{x \sim p_g}[f(x)] \right) \qquad (5)$$

WGAN uses Critical as a differentiator from the discriminator in GAN. The differences between the two are:

The final layer of Critical removes sigmoid.

There is no log item in the target function of Critical.

To ensure Lipschitz limits, Critical needs to truncate the parameters to a certain range after each update.

The better the Critical training is, the better the generator will be.

3. LSTM-GAN

In LSTM-GAN anomaly detection model, the generator aims to generate a false data distribution similar to the real sample to fool the discriminator. The generator contains three layers of LSTM. The structure of the generator is shown in the figure:
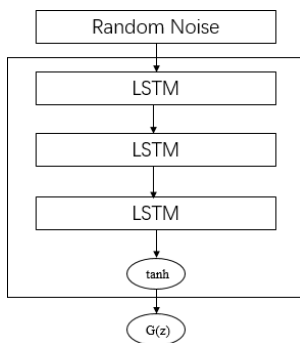

Figure 2. The structure of Generator

The discriminator needs to distinguish between real samples and false samples. Since it is necessary to fit Wasserstein distance, the objective function is:

$$\min_{G} \max_{D} V(C,G) = E_{x \sim p_{data}(x)}(C(x)) + E_{z \sim p_{z(z)}}(1 - D(G(z))) \qquad (6)$$

4. Model Evaluation

In this paper, real world datasets from NAB dataset are used in the experiment. Since F1 can give consideration to both the accuracy and recall of the model, it serves as the main indicator of the evaluation model. The experimental results are as follows：

<p align="center">**Table 1.** Experimental results</p>

Methods	Art	AdEx	AWS	Traf	Tweets	Mean
LSTM	0.375	0.538	0.474	0.643	0.543	0.509
Arima	0.353	0.583	0.518	0.571	0.567	0.524
LSTM-GAN	0.545	0.632	0.676	0.684	0.678	0.643

Bold font is the highest value. It can be seen from the table that the model proposed in this paper performs best on each dataset.

5. Conclusion

This paper proposes LSTM-GAN model for anomaly detection of time series data. This model uses LSTM as the basic network of generator and discriminator, and uses Wasserstein distance to replace the original measurement method. In order to verify the performance of the model, the model is tested on several real-world datasets, and the model proposed in this paper has good anomaly detection results.

References

[1] C. -K. Lee, Y. -J. Cheon and W. -Y. Hwang, "Studies on the GAN-Based Anomaly Detection Methods for the Time Series Data," in IEEE Access, vol. 9, pp. 73201-73215, 2021, doi: 10.1109/ACCESS.2021.3078553.

[2] Su Y , Zhao Y , Niu C , et al. Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network[C]// the 25th ACM SIGKDD International Conference. ACM, 2019.

[3] Tang J , Chen Z , Fu A , et al. Enhancing Effectiveness of Outlier Detections for Low Density Patterns[J]. Springer-Verlag, 2002.

[4] MA Biao, JIA Jun-cheng, DONG Guo-zhu et al. WGAN: Industrial Control Sensor Data Anomaly Detection Method Based on Wavelet Transform and Attention Mechanism. [J/OL]. Journal of Chinese Computer Systems,2021:1-12.

[5] Malhotra P , Vig L , Shroff G , et al. Long Short Term Memory Networks for Anomaly Detection in Time Series[C]// 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2015. 2015.

[6] R. Hsieh, J. Chou and C. Ho, "Unsupervised Online Anomaly Detection on Multivariate Sensing Time Series Data for Smart Manufacturing," 2019 IEEE 12th Conference on Service-Oriented Computing and Applications (SOCA), 2019, pp. 90-97, doi: 10.1109/SOCA.2019.00021.

[7] Goodfellow I J, Pouget-Abadie J, Mirza, et al. Generative Adversarial Networks[J]. Advances in Neural Information Processing Systems, 2014, 3:2672-2680.

[8] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN[J]. arXiv preprint arXiv:1701.07875, 2017.M.