A New Time-Series Anomaly Detection Model in CAVs Based on WGAN

Jian Yin[1], JianJun Zeng[2], and ZhenJiang Zhang[1(✉)]

[1] College of Electronics and Information Engineering, Beijing Jiaotong University, Beijing, China
zhangzhenjiang@bjtu.edu.cn
[2] College of Intelligence and Computing, Tianjin University, Tianjin, China

Abstract. As a new kind of intelligent vehicle, Connected and Automated vehicles (CAVs) can provide more convenient service by exchanging data with other vehicles and roadside units. However, CAVs rely heavily on the sensor data and communicated information. Negative factors such as faults, e1rrors, or network attacks may lead to serious consequences. To avoid the aforementioned situation, this paper proposes a method for time-series anomaly detection in CAVs. Firstly, we process the data by performing a moving standard deviation and maximum minimum normalization transformation. Then, a WGAN network with gradient constraints is constructed to perform real-time anomaly detection. We conducted experiments on the Safety Pilot Model Deployment (SPMN) dataset. Results have shown that this method can effectively improve the accuracy and sensitivity of detecting different types of abnormal situations. It also works well in detecting less obvious abnormal data.

Keywords: Connected and Automated vehicles (CAVs) · anomaly detection · WGAN

1 Introduction

Since Connected and Automated Vehicles (CAVs) have revolutionized the industry of transportation and have brought excellent benefits in decreasing the likelihood of accidents, improving quality of life, and leveraging the efficiency of Intelligent Transportation Systems (ITS) [1], most researchers have focused on this area. CAVs use wireless technology to communicate with other vehicles, roadside units (RSU) and personal mobile devices. The data sharing, such as real-time traffic and weather conditions, will meet the information need for drivers. Meanwhile, these communications will provide environmental trans-portability, and security developments that current communication systems are not capable of providing [2].

Although the increasing use of CAVs leads to many advantages, potential drawbacks cannot be ignored. The effectiveness of CAVs heavily relies on the security, precision, and reliability of sensor readings [3]. If the sensor data are used directly without anomaly detection, it may lead to serious accidents. Therefore, it is crucial to determine whether

D.-J. Deng and J.-C. Chen (Eds.): SGIoT 2024, LNICST 633, pp. 19–25, 2026.
https://doi.org/10.1007/978-3-031-93825-2_3

there are vulnerabilities in the time-series data transmitted by CAVs, and it is expected that real-time detection can be achieved through technical means.

Anomaly detection can be used to find special results deviating significantly from other observed ones and has been used in various areas, such as the industrial area, the medical area, the energy area and so on. We focus on the methods in CAVs. It has been learnt that Wyk et al. proposed an anomaly detection method in CAVs using a hybrid method of CNN-KF-$\chi 2$ -detector [1]. Javed et al. presented an anomaly detection method in CAVs using a hybrid multi-stage attention method with CNN-LSTM [4]. Moradi et al. proposed a hybrid method of Yager rules to resolve conflicts between data from different sensors to avoid hacker injection attacks [5].

Inspired by existing research, We proposed a time-series anomaly detection model in CAVs based on WGAN network and also a gradient constraint strategy is used to improve the stability of the model. Our main contributions are as follows:

- We proposed a method for data preprocessing using moving standard deviation (MSD) and min-max normalization methods jointly. Sliding Window method is used to obtain the final input data sequence.
- We proposed a time-series anomaly detection method based on WGAN with gradient constraint strategy. It can ensure the vehicles to obtain sensor data stably.
- We validated the method proposed in this paper on the Safe Pilot Model Deployment (SPMD) dataset, an open dataset provided by the US government.

The remainder of this paper is detailed as follows. Section 2 illustrates our anomaly detection method. Section 3 introduces the dataset used in the experiment and analyses the experimental results. Finally, Sect. 4 summarizes the work of this paper and discuss the future goals.

2 Method

In this section, we introduce the method of data preprocessing and the fundamental principles of WGAN network. Based on this, we introduce the network frame for time-series anomaly detection in CAVs with gradient constraint strategy.

2.1 Data Preprocessing

We map original data series into another new one through passing each raw data to a mathematical action, which typically displays useful attributes for prediction. The new data series will be normalized subsequently. Finally, a sliding window will be used to obtain a data series which can be input into the network.

Moving Standard Deviation (MSD). Moving standard deviation (MSD) method works as a statistical function in our research, which performs well during the experiment. The sequence S is transformed into $S^{(k)}$ through the method, where the standard deviation of k successive values of S is every value in $S^{(k)}$. The moving standard deviation with the window size equal to k can be calculated as:

$$S^{(k)}[i] = \sqrt{\frac{\sum_{j=i}^{i+k-1}\left(X_j - \overline{X}\right)^2}{k-1}}, 1 \le i \le n-k+1 \tag{1}$$

where X_j represents the values of original series and $\overline{X}$ indicates the mean of all the values included in the window of size k. In our experiment, the size of k is set as 5.

Min-Max Normalization. Min-max normalization is the process of scaling all values in the series to a consistent range, which helps avoid the negative influence on the neural network training from the excessively large or small data. It can be calculated as:

$$f' = \frac{f - \min(F)}{\max(F) - \min(F)} \tag{2}$$

where f represents the original data, and F illustrates the data series. f' denotes the normalized data.

Sliding Windows. The sliding window which size is 10 is used to group the processed data from 3 sensors into a matrix with size of 10 * 3 in this paper.

2.2 Wgan

WGAN is an improved network of GAN, consisting of two basic parts: a generator network (represented as G) and a discriminator network (represented as D). With constantly interacting and optimizing between the two networks, the abnormal data can be found under given rules. The structure of WGAN network is shown as Fig. 1.

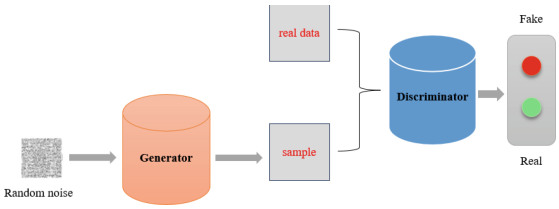

Fig. 1. The structure of WGAN network

Generator. The main function of the generator in WGAN network is to generate a set of sample values similar to real values (Random noise vectors are used as initial data in this paper.), and gradually decrease the difference from the real values. It is expected that the discriminator will be deceived until it can not distinguish the samples from the real values.

Discriminator. The core of WGAN lies in using Wasserstein distance instead of Jensen Shannon (JS) divergence as the loss function of the discriminator [6]. The Wasserstein distance can be calculated as:

$$W(P_r, P_g) = inf_{\gamma \sim \prod(P_r, P_j)} E_{(x,y)\sim\gamma}\left[||x - y||\right] \tag{3}$$

By constructing a discriminator which satisfies the K-Lipschitz condition and maximizing its estimation of Wasserstein distance between real and generated samples, the

training of the generator can be ensured to converge to the globally optimal solution. The loss function of the discriminator under the condition of fixed generator with gradient constraint can be calculated as:

$$\text{Loss}(G, D) = E_{z\sim P_z}\left[D(G(z)] - E_{x\sim P_{data}}[D(x)] + \lambda E_{x\sim P_x}\left[||\nabla_x D(x)||_2 - 1 \quad (4)\right.\right.$$

where $D(x)$ represents the scores for real data, $D(G(z))$ illustrates the scores for generated data $G(z)$. The loss is expected to be maximized in order to widen the Wasserstein distance between real data and generated data.

According to aforementioned principles, the time-series anomaly detection in CAVs can be carried out. The framework diagram is shown in Fig. 2.

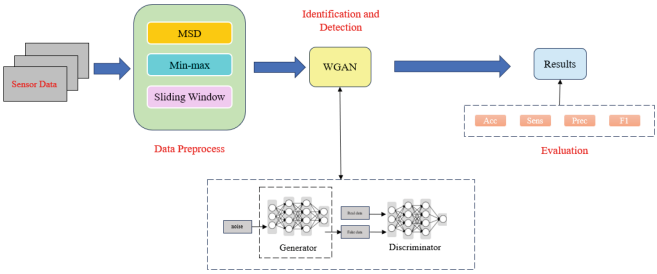

Fig. 2. The diagram of anomaly detection framework

3 Experiments

3.1 Dataset

The data for this study is obtained from the research data exchange (RDE) database for the Safe Pilot Model Deployment (SPMD) program [7]. The raw data in the SPMD dataset does not contain outliers, so we simulated a group of abnormal situations according to reference [1]. The abnormal values are mainly Gaussian noise amplified by a certain factor used to simulate the situations as follows.

- Instant: A sharp, unexplained change in the observed data between two successive sensor readings.
- Constant: A temporarily constant observation different from the "normal" sensor readings and not related to underlying physical phenomena.
- Gradual drift: A small and gradual drift in observed data during a time period.
- Bias: A temporarily constant offset from the sensor readings.

3.2 Results and Analysis

It is assumed that there is only one type of anomaly at the same time. The performance of our model will be evaluated as follows while injecting different types of anomalies in terms of accuracy, sensitivity, precision, and F1 score. The anomaly rate is set as 5%.

Instant Anomaly. The Gaussian random noise amplified by 25, 100,500,1000 and 10000 times is added as anomaly respectively. The evaluation results of MSALSTM, WKN-OC and our method are displayed in Table 1.

Table 1. Detection performance of instant anomaly type

MSALSTM-CNN(%)						WKN-OC(%)				WGAN(%)			
Magnitude	Duration	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1
U (0,1)	10	93.0	90.8	98.7	94.6	98.4	94.3	99.0	96.7	99.0	97.1	98.9	98.0
U (0,3)	10	96.4	95.4	98.9	97.2	99.2	97.8	99.5	98.6	99.4	98.0	99.1	98.5
U (0,5)	10	96.6	95.6	99.3	97.4	99.7	98.8	99.1	99.0	99.7	98.2	99.6	99.0
U (0,5)	5	95.4	92.3	99.0	95.5	99.0	97.3	99.0	98.1	99.6	98.2	99.6	99.0
U (0,5)	3	95.5	90.3	99.5	94.7	98.9	97.1	99.0	98.0	99.3	97.7	99.4	98.5

Constant Anomaly. We add constant anomaly data with different durations and standard deviations to the raw data, evaluate and compare the performance of these three methods. The results are displayed in Table 2.

Table 2. Detection performance of constant anomaly type

	MSALSTM-CNN(%)				WKN-OC(%)				WGAN(%)			
Magnitude	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1
25*N (0,0.01)	84.1	54.6	98.1	70.2	96.5	65.8	96.1	72.9	98.2	72.5	96.7	82.9
100*N (0,0.01)	95.8	89.6	98.4	93.8	99.0	90.7	98.8	94.4	99.1	92.2	97.6	94.8
500*N (0,0.01)	96.0	99.8	97.8	99.0	99.9	99.2	99.8	99.4	99.6	99.1	99.4	99.2
1000*N (0,0.01)	99.0	98.2	99.2	98.7	99.9	99.3	99.9	99.6	99.9	99.5	99.9	99.7
10000*N (0,0.01)	99.4	98.9	99.8	99.3	99.9	99.8	99.9	99.9	99.9	99.9	99.9	99.9

Gradual Drift Anomaly. Abnormal data with gradual drift may gradually move away from normal values over time. We add a set of linear increments to the normal data to simulate this abnormal situation. The results are displayed in Table 3.

Bias Anomaly. Bias outliners can be obtained by adding or subtracting a fixed value from normal data. The increase in offset amplitude and duration both contribute to the improvement of detection performance. The results are displayed in Table 4.

Table 3. Detection performance of gradual drift anomaly type

MSALSTM-CNN(%)						WKN-OC(%)				WGAN(%)			
Magnitude	Duration	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1
line space (0,4)	10	96.0	95.6	99.1	97.5	98.3	97.3	98.7	98.0	98.5	97.2	99.2	98.2
line space (0,4)	20	96.2	96.0	99.3	97.6	98.5	97.7	98.8	98.3	99.2	98.1	99.1	98.6
line space (0,2)	10	94.4	93.1	99.1	95.6	98.0	97.0	98.5	97.7	98.6	97.1	98.8	97.9
line space (0,2)	20	94.1	94.1	99.5	96.0	97.7	96.3	98.2	97.2	98.7	97.5	99.0	98.2

Table 4. Detection performance of bias anomaly type

MSALSTM-CNN(%)						WKN-OC(%)				WGAN(%)			
Magnitude	Duration	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1
U (0,1)	10	88.6	85.8	95.9	90.6	98.4	94.3	99.0	96.7	99.0	97.2	99.3	98.2
U (0,3)	10	95.0	93.8	98.5	96.1	99.2	97.8	99.5	98.6	99.7	97.5	99.4	98.1
U (0,5)	10	96.6	95.7	99.1	97.4	99.7	98.8	99.1	99.0	99.8	98.0	99.5	98.7
U (0,5)	5	95.5	91.2	99.6	95.6	99.0	97.3	99.0	98.1	99.6	97.8	99.2	98.5
U(0,5)	3	95.1	90.5	99.5	94.8	98.9	97.1	99.0	98.0	99.2	97.6	99.0	98.3

The above experimental results indicate that our method has good performance in detecting abnormal data. Even if the duration is short and the amplitude of abnormal data is small, it still has stable performance, which is significantly improved compared to MSALSTM-CNN and WKN-OC methods.

4 Conclusion

The main objective of this study is to improve the security of CAVs by proposing a time-series anomaly detection method. The WGAN network with gradient constraints can effectively identify different types of anomalies, maintaining good performance throughout the experimental process and improving the score of anomaly detection. In future research, emphasis will be placed on how to reduce computation time and resource requirements while maintaining high detection rates, making it more suitable for practical deployment on CAVs.

Acknowledgment. This work was supported by the National Natural Science Foundation of China under Grant 62173026.

References

1. van Wyk, F., Wang, Y., Khojandi, A., Masoud, N.: Real-time sensor anomaly detection and identification in automated vehicles. IEEE Trans. Intell. Transp. Syst. **21**(3), 1264–1276 (.2020). https://doi.org/10.1109/TITS.2019.2906038
2. Khanmohammadi, F., Azmi, R.: Time-series anomaly detection in automated vehicles using D-CNN-LSTM autoencoder. IEEE Trans. Intell. Transp. Syst. https://doi.org/10.1109/TITS.2024.3380263
3. Khan, I., Moustafa, N., Pi, D., Haider, W., Li, B., Jolfaei, A.: An enhanced multi-stage deep learning framework for detecting malicious activities from autonomous vehicles. IEEE Trans. Intell. Transp. Syst. **23**(12), 25469–25478 (2022)
4. Javed, A.R., Usman, M., Rehman, S.U., Khan, M.U., Haghighi, M.S.: Anomaly detection in automated vehicles using multistage attention-based convolutional neural network. IEEE Trans. Intell. Transp. Syst. **22**(7), 4291–4300 (2021). https://doi.org/10.1109/TITS.2020.3025875
5. Moradi, M., et al.: Sensor and decision fusion-based intrusion detection and mitigation approach for connected autonomous vehicles. IEEE Sens. J. **24**(13), 20908–20919 (2024)
6. Arjovsky, M., Chintala, S., Léon, B.: Wasserstein GAN (2017). 10.48550/ arXiv. 1701.07875
7. Bezzina, D., Sayer, J.: Safety pilot model deployment: test conductor team report. U.S. Dept. Transp., Washington, DC, USA, Tech. Rep. DOT HS 812, 2014