2024 4th International Conference on Neural Networks, Information and Communication (NNICE) | 979-8-3503-9437-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/NNICE61279.2024.10498833

An Attack Prediction and Recognition Method for Water Treatment System Based on One-dimensional Convolutional Neural Network and Cumulative Sum

Xiangdong Hu*
College of Automation/College of Industrial Internet
Chongqing University of Posts and Telecoms.
Chongqing, China
*Corresponding author: huxd@cqupt.edu.cn

Lang Liu
College of Automation/College of Industrial Internet
Chongqing University of Posts and Telecoms.
Chongqing, China
liul1652725@163.com

***Abstract*-An attack prediction and recognition method for water treatment system based on one-dimensional convolutional neural network (1D-CNN) and cumulative sum (CUSUM) is proposed in this paper. Firstly, a 1D-CNN is employed as a time series predictor to establish an independent prediction model for each attack point, followed by training and learning. Subsequently, the deviation between predicted values and actual values is compared with a threshold value using CUSUM statistics to identify anomalies within the water treatment system. Experimental evaluations conducted on the Safe Water Treatment (SWaT) Test Bench demonstrate that 1D-CNN exhibits superior predictive performance compared to other models. Furthermore, CUSUM successfully detects and identifies all listed attacks based on statistical bias.**

Keywords-water treatment system; 1D-CNN; CUSUM; prediction; recognition

I. INTRODUCTION

In recent years, the industrial control system has continuously integrated new technologies and advanced network communication, giving rise to a new generation of industrial Internet systems [1]. As an integral component of urban infrastructure, water treatment systems play a pivotal role in sustaining urban life and supporting industrial activities [2]. However, with numerous devices connected to the network and operating online simultaneously, coupled with inadequate comprehensive protection mechanisms, the attack surface continues to expand. Consequently, ensuring the secure operation of water treatment systems poses significant security risks. Therefore, it is imperative to develop an effective method for detecting attacks in order to guarantee the safe operation of water treatment systems.

Anomaly detection is an effective method for identifying abnormal events in a dataset that deviate significantly from normal patterns. Previous studies [3-5] have utilized normal datasets from water treatment systems for training and validation, enhancing the model's ability to detect attacks while considering computational requirements. However, these methods primarily focus on post-attack detection, necessitating early and accurate identification of anomalies in critical infrastructure like water treatment systems to prevent system damage and service disruption. In contrast, literature [6] employed multi-layer perceptron (MLP), while literature [7] used 1D-CNN for prediction; however, no comparison or analysis with other prediction models was conducted to verify their superior performance or address real-time requirements.

The present paper proposes a prediction and recognition method for water treatment systems based on 1D-CNN and CUSUM to address these issues. CUSUM calculates the deviation between the predicted output of 1D-CNN and the actual value, comparing it against upper and lower control limit thresholds for detection and identification purposes. Comparative analysis with LSTM, RNN, GRU, and other five prediction models demonstrates significant improvement in prediction performance using the proposed method, along with notable advantages in training time and prediction time index.

II. PREDICTION AND RECOGNITION METHOD OF WATER TREATMENT SYSTEM

A. Framework of the model

In this study, we employ a 1D-CNN model to accurately forecast the future attack point values in water treatment systems. Additionally, CUSUM analysis is utilized to monitor the cumulative deviation between predicted and actual values. The process of predicting and identifying attack points in the water treatment system based on 1D-CNN and CUSUM is illustrated in Fig. 1.

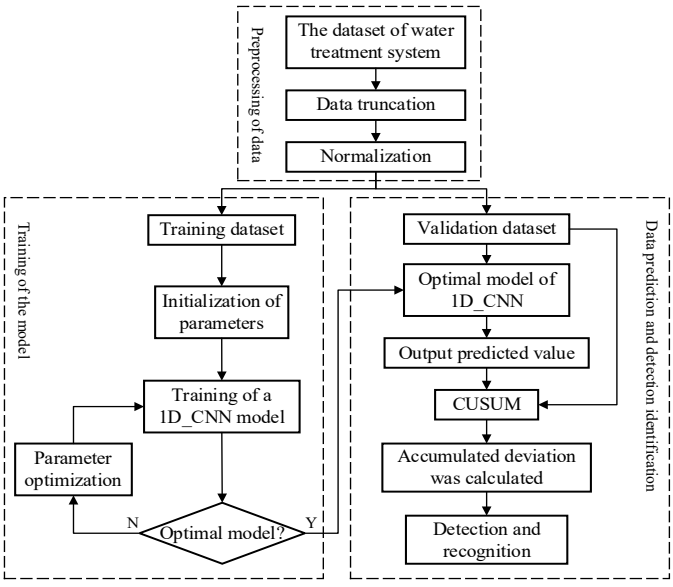

Figure 1. Prediction and recognition process of water treatment system

The data preprocessing component performs tailored and normalized processing on the water treatment system dataset to

facilitate subsequent training and identification tasks. For model training, each attack point is divided into independent 1D-CNNs for individualized learning and prediction transfer to the next stage for detection and recognition purposes. In the detection and recognition phase, statistical deviation between predicted values and actual values is compared using CUSUM analysis to detect deviations exceeding a predefined threshold.

B. Principle of 1D_CNN

Given the one-dimensional nature of data pertaining to a single attack point in water treatment systems, and considering the real-time demands of industrial control systems akin to water treatment systems, this study employs 1D-CNN[8] for predictive analysis.

The 1D-CNN convolutional layer formula is as follows:

$$c_i^{m+1}(j) = K_i^m * x^m(j) + b_i^m \tag{1}$$

Where "*" represents the convolution of the filter kernel and the local region; $c_i^{m+1}(j)$ represents the input of the j neuron in the i channel of layer $m+1$; K_i^m represents the weight of the i filter kernel in the m layer; b_i^m represents the bias corresponding to the i filter kernel in the m layer.

After the convolution operation, the activation function is employed to acquire the nonlinear representation of the input signal. In this study, a modified rectified linear unit (ReLU) is utilized as the activation function for 1D-CNN. The specific formulation is presented as follows:

$$\alpha_i^{m+1}(j) = f(c_i^{m+1}(j)) = max\{0, c_i^{m+1}(j)\} \tag{2}$$

Where $c_i^{m+1}(j)$ represents the output value of the convolution operation, and the output value $\alpha_i^{m+1}(j)$ after input to the activation function is.

C. C. Principle of CUSUM

In this study, the CUSUM method is employed to monitor the cumulative sum of continuous observations, when the accumulation surpasses this threshold, it indicates a significant process change, triggering timely alarms and appropriate actions [9]. The detailed explanation is provided below.

Hypothesis $x_1, x_2, \cdots, x_\zeta, \cdots, x_n$ is a set of process monitoring data recorded in sequence. The following assumptions are made for this set of data, expressed by equation (3) :

$$\begin{aligned} &H_0: \ x_i \sim N(\mu_0, \sigma), i = 1,2,3,...,n \\ &H_1: \ x_i \sim N(\mu_0, \sigma), i = 1,2,3,...,\zeta \\ &x_i \sim N(\mu_1, \sigma), i = \zeta+1, \zeta+2,...,n \end{aligned} \tag{3}$$

Where, H_0 is the null hypothesis, indicating that the process has no change point; H_1 is the alternative hypothesis, indicating that the process has a change point. μ_0, μ_1 are the sample mean before and after the occurrence of the change point respectively, and the change point is the mutation point of the sample data, corresponding to the attack point in the water treatment system. σ is the sample standard deviation; n is the upper limit of the sample data group; ζ is the data group where the change point occurs, and $\zeta < n$. The likelihood ratio of H_1 to H_0 can be expressed by equation (4) :

$$L_{n,\zeta} = \frac{L(n,\mu_1)}{L(n,\mu_0)} = \frac{\prod_{i=1}^{\zeta} f_0(x_i) \prod_{i=\zeta+1}^{n} f_1(x_i)}{\prod_{i=1}^{n} f_0(x_i)} = \frac{\prod_{i=\zeta+1}^{n} f_1(x_i)}{\prod_{i=\zeta+1}^{n} f(x_i)} \tag{4}$$

Where, $L_{n,\zeta}$ is the likelihood ratio statistic of alternative hypothesis H_1 to null hypothesis H_0 ; $L(n,\mu_0)$、$L(n,\mu_1)$ are likelihood functions of H_0 and H_1 , respectively; $f_0(x_i)$、$f_1(x_i)$ are density functions of H_0 and H_1 , respectively.

III. EXERIMENTAL RESULTS AND ANALYSIS

A. Introduction to the dataset

The SWaT dataset adopted in this paper was developed by iTrust Network Security Research Center of Singapore University of Technology and Design [10]. The test bed consists of six processes from stage 1 to Stage 6, which work together to treat water resources in order to simulate the treatment process of a large urban water plant.

SWaT dataset contains 25 sensors and 26 actuators. In order to verify the effectiveness of the proposed method, this paper only studies sensors and actuators in stage 1 of SWaT dataset, namely FIT101, LIT101, MV101 and P101. They are attack points that can be attacked in stage 1.

B. Experimental environment and evaluation index

The algorithm test experiments in this paper were carried out under the operating system Windows10, processor Intel(R) Core(TM) i7-10700 CPU @ 2.90GHz, and 16.0GB RAM. The algorithm implementation of the model calls TensorFlow, Scikit-learn, Keras library and other methods in Python.

The method proposed in this paper uses 1D-CNN to predict the attack point value in water treatment system, and the accuracy and real-time performance of the model prediction are very important. For accuracy, the evaluation indexes of experimental selection include Mean Absolute Error (*MAE*), Root Mean Square Error (*RMSE*), Coefficient of Determination(R^2). The formulas are shown below.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - y_i^{'} \right| \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y_i^{'})^2} \tag{6}$$

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i' - \bar{y})^2} \quad (7)$$

Where, n is the number of samples; y_i is the actual value of the i sample; y_i' is the predicted value of the i sample; $\bar{y}$ is the sample mean.

C. Experimental results

1) Prediction result analysis

FIT101, LIT101, MV101 and P101 were used to compare and experiment with five prediction models. The comparison model includes long short-term memory (LSTM), recurrent neural network (RNN), gate recurrent unit (GRU), bi-directional long short-term memory (BiLSTM) and bi-directional gated recurrent unit (BiGRU). The experimental results are shown in Table I.

In terms of MAE, except that RNN based on data P101 is better than the model in this paper, the MAE index of the model in this paper is optimal for the other three groups of data. For data LIT101, the MAE value of the model in this paper is at least 0.0025, which is more than 10 times lower than that of the five comparison models, indicating that the prediction error of the model in this paper is smaller and more accurate. The situation of $RMSE$ is similar to that of MAE. Although the RMSE value of the model in this paper is not optimal for MV101 and P101, it is also very close to the optimal value. For R^2, the R^2 value of the model in this paper for LIT101 data reaches 0.9999, which is very close to 1, indicating that the representation integration between the model in this paper and LIT101 data is excellent.

TABLE I. Comparison results of prediction experiments

Criterion for evaluation	The name of the data	Predictive modeling framework					
		LSTM	RNN	GRU	BiLSTM	BiGRU	Proposed method
MAE	FIT101	0.0108	0.0085	0.0073	0.0060	0.0082	0.0048
	LIT101	0.0544	0.0488	0.0632	0.0540	0.0579	0.0025
	MV101	0.0088	0.0076	0.0162	0.0051	0.0085	0.0041
	P101	0.0050	0.0030	0.0061	0.0047	0.0049	0.0046
RMSE	FIT101	0.0191	0.0205	0.0188	0.0169	0.0170	0.0139
	LIT101	0.0837	0.0738	0.0914	0.0685	0.0755	0.0032
	MV101	0.0439	0.0439	0.0455	0.0437	0.0438	0.0438
	P101	0.0494	0.0495	0.0495	0.0495	0.0495	0.0496
R^2	FIT101	0.9979	0.9976	0.9980	0.9985	0.9983	0.9989
	LIT101	0.9393	0.9528	0.9275	0.9594	0.9506	0.9999
	MV101	0.9634	0.9634	0.9608	0.9638	0.9635	0.9636
	P101	0.9870	0.9869	0.9869	0.9869	0.9869	0.9868

2) Time contrast analysis

Six forecasting models were used to train and predict the four groups of data respectively. The training time of each forecasting model on the four groups of data and the time used to predict a single sample were averaged, and the results of the training time and prediction time were shown in Table II.

TABLE II. Results of training time and predicted time

Prediction model	Average training time/second	Average time to predict a single sample/second
LSTM	390.7175	0.0454
RNN	257.4748	0.0372
GRU	441.0330	0.0482
BiLSTM	455.0847	0.0549
BiGRU	525.4947	0.0481
Proposed model	155.7058	0.0196

It can be seen from Table II that the average training time and the average time for predicting a single sample of the model in this paper are both the shortest, which verifies the real-time and effectiveness of the model in predicting the attack point of the water treatment system.

3) Detection and recognition

The upper and lower control limits of FIT101, LIT101, MV101 and P101 were obtained by CUSUM using validation set data, as shown in Fig. 2. The upper and lower control limits are used as thresholds to compare with subsequent accumulations and deviations.

Since the model in this paper is only trained to simulate the normal behavior of stage 1 in the SWaT test bench, we only consider the attacks in stage 1, and the attack description is shown in Table III. A total of 4 attacks were detected through

the cumulative deviation between the predicted value and the actual value of CUSUM statistics, and all the attacks in Table III were detected. The distribution of CUSUM values for sensors LIT101 and P101 is illustrated in Fig. 3 and Fig. 4.

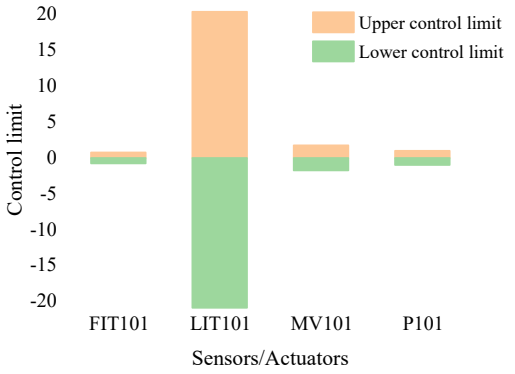

Figure 2. Upper and lower control limits

TABLE III. Attack description in phase 1

Attack ID	Attacker's intent	Description of attack
A1	Overflow tank	Open MV-101
A2	Underflow the tank and damage P-101	Increase water level by 1 mm every second
A3	Overflow tank	Keep MV-101 on continuously; Value of LIT-101 set as 700 mm
A4	Underflow the tank and damage P-101	Set LIT-101 to above H threshold

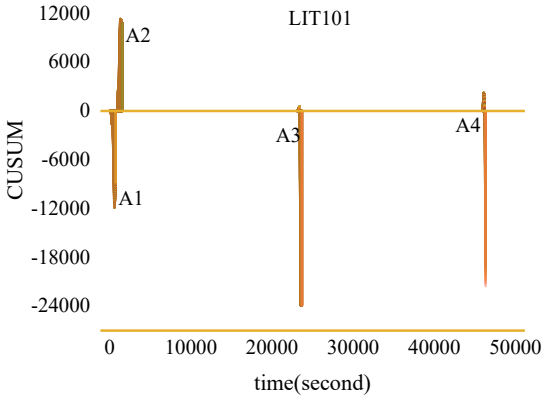

Figure 3. Revealed incidents of attacks in LIT101

IV. CONCLUSIONS

This paper proposes a prediction and identification method based on 1D-CNN and CUSUM. Firstly, a prediction network using 1D-CNN is constructed for each attack point, followed by training and learning to make predictions. Subsequently, the predicted values are analyzed alongside actual values using CUSUM to compare against upper and lower control limit thresholds for detection and identification. The 1D-CNN prediction network exhibits superior predictive performance compared to LSTM, RNN, GRU, BiLSTM, and BiGRU models. Furthermore, all four types of attacks mentioned in this paper are accurately identified.

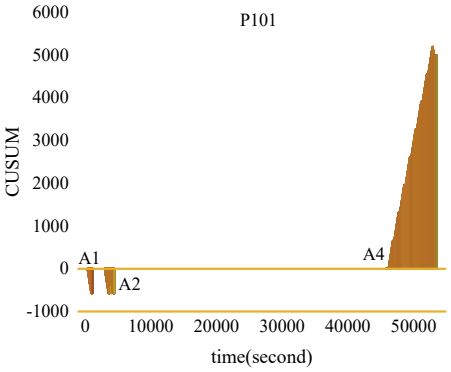

Figure 4. Revealed incidents of attacks in P101

References

[1] Jiang, Y., Wu, S., Ma, R., Liu, M., Luo, H., & Kaynak, O. (2023). Monitoring and defense of industrial cyber-physical systems under typical attacks: From a systems and control perspective. IEEE Transactions on Industrial Cyber-Physical Systems.

[2] Alimi, O. A., Ouahada, K., Abu-Mahfouz, A. M., Rimer, S., & Alimi, K. O. A. (2021). A review of research works on supervised learning algorithms for SCADA intrusion detection and classification. Sustainability, 13(17), 9597.

[3] Elnour, M., Meskin, N., Khan, K., & Jain, R. (2020). A dual-isolation-forests-based attack detection framework for industrial control systems. IEEE Access, 8, 36639-36651.

[4] Elnour, M., Meskin, N., & Khan, K. M. (2020, August). Hybrid attack detection framework for industrial control systems using 1D-convolutional neural network and isolation forest. In 2020 IEEE Conference on Control Technology and Applications (CCTA) (pp. 877-884). IEEE.

[5] Abdelaty, M., Doriguzzi-Corin, R., & Siracusa, D. (2021). DAICS: A deep learning solution for anomaly detection in industrial control systems. IEEE Transactions on Emerging Topics in Computing, 10(2), 1117-1129.

[6] MR, G. R., Somu, N., & Mathur, A. P. (2020). A multilayer perceptron model for anomaly detection in water treatment plants. International Journal of Critical Infrastructure Protection, 31, 100393.

[7] Kravchik, M., & Shabtai, A. (2018, January). Detecting cyber attacks in industrial control systems using convolutional neural networks. In Proceedings of the 2018 workshop on cyber-physical systems security and privacy (pp. 72-83).

[8] Katranji, A., Shafiullah, M., & Rehman, S. (2023, October). Short-Term Wind Speed Prediction for Saudi Arabia via 1D-CNN. In 2023 IEEE 13th International Conference on System Engineering and Technology (ICSET) (pp. 153-158). IEEE.

[9] Bojović, P. D., Bašičević, I., Ocovaj, S., & Popović, M. (2019). A practical approach to detection of distributed denial-of-service attacks using a hybrid detection method. Computers & Electrical Engineering, 73, 84-96.

[10] Goh, J., Adepu, S., Junejo, K. N., & Mathur, A. (2017). A dataset to support research in the design of secure water treatment systems. In Critical Information Infrastructures Security: 11th International Conference, CRITIS 2016, Paris, France, October 10–12, 2016, Revised Selected Papers 11 (pp. 88-99). Springer International Publishing.