False data injection attacks detection on power systems with convolutional neural network

Min Lu¹, Lijian Wang¹, Zhangjie Cao¹, Youguo Zhao² and Xiangyang Sui²

¹State Grid Zhejiang Electric Power Company, Hangzhou 310009, China
²Dongfang Electronics Co., Ltd, Yantai 264000, China

E-mail: 482886958@qq.com

Abstract. False data injection attack (FDIA) could manipulate measurement information collected by supervisory control and data acquisition (SCADA) system, which tempts with decisions of power grid and threatens the state estimation of smart grid. Aiming at the state estimation of smart grid, the principle of FDIA under the AC power flow model is studied and a FDIA detection model based on improved convolutional neural network CNN is constructed. By adding the gate recurrent unit (GRU) to the fully connected layer in CNN, the CNN-GRU network is designed to train and update network parameters based on historical measurement data of power grid, and extract spatial and temporal characteristics of the data to implement efficient and real-time FDIA detector. Finally, in the IEEE 118 bus test systems, experiments are carried out to verify the effectiveness of the proposed method.

1. Introduction
With the advancement of network communication technology[1-4], the operational reliability of the power system is significantly enhanced, but due to the dependence of the smart grid on data communication, it is vulnerable to a variety of malicious network attacks, among which false data injection attack (FDIA) can tamper with the measurement information collected by the data acquisition and monitoring (SCADA) system, affecting the important decisions of the grid, thus posing a security threat to the smart grid state estimation. How to detect FDIA in real time and efficiently for the protection of the smart grid safe operation is of great significance.

A number of detection methods have been proposed in recent years for different FDIA. Common methods are based on graph theory and distance to detect FDIA [5-6]. Among them, the Literature [1] proposed a distributed malicious data injection detection method based on bus phase angle Markov graph, using the measured value equivalent conversion and the maximum weighted residual model to jointly detect malicious data injection attacks. In [6], a method for tracking dynamic measurement changes is proposed to detect FDIA, and the KL distance (Kullback-Leibler divergence) is used to calculate the distance value of the probability distribution. However, this method does not handle continuous small-scale attacks and replay attacks well. This type of method identifies malicious data attacks by probability distribution detection and is greatly affected by historical data.

Another type of study uses sequence detection to detect FDIA [7-8]. In [7], a sequential detection method based on generalized likelihood ratio is proposed to deal with spurious data attacks, and a distributed adaptive sampling sequence detection based on hierarchical trigger sampling is further developed, which can effectively reduce communication cost and improve detection efficiency. Literature [8] proposed a real-time detection framework to detect malicious data injection attacks,
which can process unknown parameter variables in a low-complexity method and process multiple measurements at a time.

In recent years, with the development of smart grids, traditional methods are not sufficient to cope with the increase in the actual grid data volume. Machine learning and deep learning methods are increasingly applied to FDIA testing and are significantly improved over traditional methods [9-14]. Literature [10] used the machine learning method for malicious data injection attack identification in smart grid, and proposed an online batch learning algorithm to model this attack problem, and proposed a detection method based on statistical model and graph model. In [11], deep learning techniques are used to identify FDIA's behavioral characteristics using historical measurement data, and the captured features are used to detect attacks in real time. Literature [12] proposed a detection method based on supervised learning classifier, and compared direct and invisible FDIA.

Although the existing detection methods have been able to achieve the detection of FDIA, due to the complex topology information of the grid, the amount of data on the smart grid is getting larger and larger, and the detection accuracy of common detection methods is not enough. In order to improve the accuracy of FDIA detection, this paper proposes a detection method based on improved convolutional neural network, which is built by adding a gated recurrent unit (GRU) to a convolutional neural network (CNN) to construct a CNNGRU hybrid. The neural network considers the spatial and temporal characteristics of the data for feature extraction to achieve real-time and efficient FDIA detection.

2. Problem formulation

2.1. State estimation

System state estimation is a key mechanism to maintain the stability and efficiency of modern power grids [15]. In the grid, the control center needs to monitor the voltage phase angles of all buses in order to make real-time decisions on operation, but it is impractical to measure all bus voltage phase angles directly. Therefore, the control center estimates the operating state of the system by collecting measurement data from remote smart meters. The specific measurement data includes branch active power flow and bus active power injection, which can be used to estimate the bus voltage angle in the system.

Generally speaking, in the AC power flow model, the relationship between the measured values $z$ and $x$ is:

$$ z = H(x) + e $$

(1)

where, $x$ is the system state variable, the measured value for state estimation is obtained by the electricity meter and collected by the energy control center; $z$ is the measured value from the data acquisition center; $H(\cdot)$ is a nonlinear mapping of system state to measured value; $e$ is the measured value error vector.

At time $t$, the state variable of the system is $x(t)$, and the measured value is $z(t)$. The dispatching center estimates the state variable $\hat{x}(t)$ at time $t$ through the measured value $z(t)$, and the application has bad data detection. The iterative state estimation method [5] assumes that $z(t)$ satisfies the observability requirements of state estimation. Then, the state estimator presents the following weighted least squares problem to obtain $\hat{x}(t)$.

$$ \hat{x}(t) = \arg \min_{\hat{x}} (z(t) - H(x(t)))^T W (z(t) - H(x(t))) $$

(2)

where, $W$ is a known weight matrix.

After solving Equation (2), the residual of the measured value can be obtained as $z(t) - H(x(t))$. Then, the bad data detection [15] can be used to screen out suspicious measurement results before the next round of state estimation.
2.2. FDIA
Traditional FDIA is usually as shown in Equation (3).
\[ z_a = z + a = H(x) + a + e \]  \hspace{1cm} (3)
where, \( a \) is the injected false data attack vector; \( z_a \) is the damaged measure vector; \( x \) is the state estimate of the original measured value \( z \) when not attacked.

If the attacker knows the power system topology, known as the matrix function \( H(x) \), then an unobservable FDIA can be constructed [16]. Specifically, the false injection data \( a \) satisfies the Equation (4).
\[ a = H(x + e) - H(x) \]  \hspace{1cm} (4)
where, \( e \) is the false state data, \( e = [e_1, e_2, \ldots, e_n]^T \) is an arbitrary \( n \times 1 \) non-zero vector; \( n \) is the number of states; \( x_a = H^{-1}(z_a) = x + e \) is an \( n \times 1 \) vector, indicating the state estimation of the new measured value \( z_a \) under attack. Then, the damaged measurement vector \( z_a \) is expressed as follows.
\[ z_a = H(x + e) + e = H(x_a) + e \]  \hspace{1cm} (5)

The form of \( z_a \) in Equation (5) is the same as Equation (3). At this time, the residual is expressed as:
\[ r_a = z_a - H(x_a) = z + a - H(x + e) = z - H(x) \]  \hspace{1cm} (6)

Therefore, it is possible to bypass the traditional residual-based bad data detection and inject the dummy data into the system measurement value to realize FDIA.

**Figure 1.** False data injection attack detection model based on CNN-GRU network.

3. Proposed solution
Long-short-term memory (LSTM) is a special type of recurrent neural network (RNN) that can learn long-term dependence information and can solve the gradient disappearance or gradient explosion problem that may exist in RNN. As a variant of LSTM, GRU is simpler to construct than LSTM. It
saves time and converges faster when training a large amount of measured value data. Therefore, by adding the GRU to the fully connected layer in CNN, the CNN-GRU hybrid neural network is constructed, and the measured value data is trained to extract features and detect false data.

3.1. FDIA detection model based on CNN-GRU hybrid network

Traditional CNN is mostly used to extract the spatial features of images, and FDIA is closely related to the topology information of power grid structure. This paper considers the spatial characteristics and time characteristics of attack samples by CNN combined with GRU network, and constructs FDIA detection model based on CNN-GRU hybrid network. The general structure of the detection model is shown in Figure 1.

The detection model is mainly divided into two parts: the training phase and the detection phase. The training phase is mainly based on the attack samples in the historical database. The historical data is extracted by the feature extraction module of the CNN-GRU hybrid network, and the CNN-GRU network is continuously updated and adjusted according to the extraction result to obtain a suitable false injection attack signature database. The real-time data collected by the terminal is input into the trained CNN-GRU network, and the data is classified by the Softmax classifier. The data classified as abnormal will trigger the alarm module to realize FDIA detection, and if it is classified as normal, it will not alarm.

3.2. CNN-GRU network design

According to the FDIA detection model based on CNN-GRU network in Section 3.1, CNN-GRU network design and training needs to be completed before testing.

The traditional CNN is mainly composed of the following five structures.

1) Data input layer. The input layer is the input to the entire neural network. In a specific process, it is generally a multidimensional matrix.

2) Convolution layer. The convolutional layer is the most important part of CNN. Unlike the traditional connection layer, the input of each window in the convolutional layer is only a small piece of the upper layer of the neural network. The convolutional layer tries to analyze each small piece of the neural network in more depth to obtain a higher degree of abstraction. Characteristics. For a two-dimensional matrix $Z$ and a two-dimensional convolution kernel filter matrix $K$, the convolution operation can be expressed as:

$$S(i, j) = (ZK)(i, j) = \sum_{m=1}^{i} \sum_{n=1}^{j} Z(m,n)K(i-m, j-n)$$  \hspace{1cm} (7)

where, $i$ and $j$ are the horizontal and vertical axis positions of the corresponding small block of the upper layer of the neural network; $m$ and $n$ are the horizontal and vertical axis positions of the corresponding small blocks of the convolutional layer, respectively.

3) Pooling layer. The pooled layer neural network does not change the depth of the input matrix, but it can reduce the size of the matrix. Through the pooling layer, the number of nodes in the last fully connected layer can be further reduced, thereby achieving the purpose of reducing parameters in the entire neural network.

4) Fully connected layer. After multiple rounds of convolutional and pooling layers, the input information has been abstracted into features with higher information content, and the classification results are given by the fully connected layer at the end of the CNN.

5) Output layer. After the fully connected layer, CNN mainly classifies and outputs the results through the Softmax classifier.

In the CNN-GRU hybrid network proposed in this paper, before the fully connected layer, 100 GRU structures are added to process the timing characteristics of the extracted input data. The single GRU structure consists of an update gate and a reset gate.

After the CNN-GRU network structure is trained according to the above description, it can be used for FDIA detection. The steps of FDIA detection based on CNN-GRU network are as follows:
1. The original measurement data set \( \{z_i\} \) is preprocessed, and \( n \) measured value vectors \( z \) are treated as \( n \times m \) matrix \( Z \), and the expression is as follows:

\[
Z = \begin{bmatrix}
    z_{11} & \cdots & z_{1m} \\
    \vdots & \ddots & \vdots \\
    z_{n1} & \cdots & z_{nm}
\end{bmatrix}
\]  

(8)

The matrix \( Z \) is used as the input of CNN-GRU network, and then the data is de averaged and normalized.

2. The preprocessed input data is input into convolution layer, which is divided into two parts: feature extraction layer and feature mapping layer.

3. After processing through the volume accumulation layer, the results are input into the pooling layer, the input data is divided into non-overlapping areas according to the window size, and then the elements in each area are aggregated. According to the size of the data, the appropriate window size is selected for the pooling operation, so as to achieve the feature dimension reduction.

4. After several steps 2 and 3, input the results into 100 GRU structures. By updating the door and resetting the door, updating the door is used to control the extent to which the status information of the previous moment is brought into the current state. The larger the value of updating the door, the more the status information of the previous moment is brought in. Reset gate is used to control the degree of neglecting the state information of the previous moment. The more the value of reset gate is, the more it will be ignored.

5. The result is input into the last full connection layer, and the result is classified and output by Softmax classifier.

4. Case studies

The IEEE 118-bus test systems are selected as the test environment. The measured values are composed of four parts: bus phase angle, voltage amplitude, bus injected active power and reactive power, and active power and reactive power injected into each branch. A total of 3,000 measurement samples were randomly selected as experimental data. Each experimental data contained topology information of the power grid, power generation side output and power load amount, and the training and test sample sets were composed in a ratio of 2:1.

| Number | Method    | Performance | Correct | False positive |
|--------|-----------|-------------|---------|----------------|
| 1      | CNN-GRU   |             | 90%     | 5%             |
| 2      | CNN       |             | 78%     | 10%            |
| 3      | RNN       |             | 75%     | 12%            |
| 4      | DBN       |             | 71%     | 15%            |

In this paper, the proposed method is compared with other methods, and experiments are carried out on the IEEE 118-node standard test system. The experiment mainly compares the detection accuracy with different methods, i.e., the CNN-GRU method proposed in this paper, the traditional CNN-based detection method, the traditional RNN-based detection method and the deep belief network (DBN) algorithm based detection method, in the case of the same attack intensity. Considering that the attack vectors are randomly selected and injected into the normal data, this experiment tests different attack strengths. The main influencing factors of attack intensity are the sparsity of attack vector \( a \), that is, the number of damaged measurements and the variance representing the interference amplitude of false data. Therefore, the attack intensity is defined as the proportion factor of the variance of false data, which is represented by \( A \). In this experiment, the variance of
attack intensity $\sigma^2$ is limited to 0.05. Within the limited variance, the attack intensity is divided into three grades: A = 0.1 for small intensity, A = 1.0 for medium intensity, A = 10.0 for large intensity.

Table 1 shows the experimental results of the accuracy comparison of the algorithms on the IEEE 118 bus system. The experiment verifies that the accuracy of each algorithm in the IEEE 118-node system under a medium attack strength and a balanced attack data size. It can also be seen that the CNN-GRU method proposed in this paper has an accuracy rate of about 90%, a fast convergence rate, and an accuracy rate that is significantly higher than the other two methods. This shows that the proposed CNN-GRU hybrid neural network method based on this paper Performance is higher than other algorithms, including DBN-based methods, traditional RNN-based and CNN-based approaches.

5. Conclusions

1) The power system state estimation and power flow optimization under the AC model are studied, and the mechanism and method of FDIA generation are deeply studied.

2) From the perspective of detection methods, an attack detection model based on CNN-GRU hybrid neural network is constructed. This method can simultaneously extract the spatial and temporal characteristics of the original measurement data of the grid, thereby improving the detection accuracy of the existing FDIA.

3) A number of experiments were carried out in the IEEE 118-bus system. The detection accuracy of the method was tested under different attack strengths. The validity of the method was verified and compared with other algorithms on different node systems. The experimental results verify the effectiveness and advantages of the proposed method and model.

Acknowledgement

This work was supported by Technology Project of State Grid Zhejiang Electric Power Co., Ltd, Hangzhou Power Supply Company under Number 5211HZ1800P3.

References

[1] Faheem M, Shah S B H, Butt R A, et al. 2018 Smart grid communication and information technologies in the perspective of Industry 4.0: Opportunities and challenges[J]. Computer Science Review 30 1-30
[2] Faheem M, Butt R A, Raza B, et al. 2019 A multi-channel distributed routing scheme for smart grid real-time critical event monitoring applications in the perspective of Industry 4.0[J]. International Journal of Ad Hoc and Ubiquitous Computing 32(4) 236-256
[3] Faheem M, Butt R A, Raza B, et al. 2018 Bio-inspired routing protocol for WSN - based smart grid applications in the context of Industry 4.0[J]. Transactions on Emerging Telecommunications Technologies e3503
[4] Faheem M and Gungor V C 2017 Capacity and spectrum-aware communication framework for wireless sensor network-based smart grid applications[J]. Computer Standards & Interfaces 53 48-58
[5] Sedghi H and Jonckheere E 2015 Statistical structure learning to ensure data integrity in smart grid[J]. IEEE Transactions on Smart Grid 6(4) 1924-1933
[6] Gu Chaojun, Jirutitijaroen P and Motani M 2015 Detecting false data injection attacks in AC state estimation[J]. IEEE Transactions on Smart Grid 6(5) 2476-2483
[7] Li Shang, Yilmaz Y and Wang Xiaodong 2015 Quickest detection of false data injection attack in wide-area smart grids[J]. IEEE Transactions on Smart Grid 6(6) 2725-2735
[8] Huang Yi, Tang Jin, Cheng Yu, et al. 2016 Real-time detection of false data injection in smart grid networks: an adaptive CUSUM method and analysis[J]. IEEE Systems Journal 10(2) 532-543
[9] Ozay M, Esnaola I, Vural F T, et al. 2016 Machine learning methods for attack detection in the smart grid[J]. IEEE Transactions on Neural Networks and Learning Systems 27(8) 1773-1786
[10] Esmalifalak M, Liu Lanchao, Nguyen N, et al. 2017 Detecting stealthy false data injection using machine learning in smart grid [J]. IEEE Systems Journal 11(3) 1644-1652

[11] He Youbiao, Mendis G J and Wei Jin 2017 Real-time detection of false data injection attacks in smart grid: a deep learning-based intelligent mechanism[J]. IEEE Transactions on Smart Grid 8(5) 2505-2516

[12] Yan Jun, Tang Bo and He Haibo 2016 Detection of false data attacks in smart grid with supervised learning [C]. International Joint Conference on Neural Networks July 25-29, 2016, Vancouver, Canada 1395-1402

[13] Wang Huaizhi, Ruan Jiaqi, Wang Guibin, et al. 2018 Deep learning-based interval state estimation of AC smart grids against sparse cyber attacks[J]. IEEE Transactions on Industrial Informatics 14(11) 4766-4778

[14] Yan Jun, He Haibo, Zhong Xiangnan, et al. 2017 Q-learning based vulnerability analysis of smart grid against sequential topology attacks[J]. IEEE Transactions on Information Forensics and Security 12(1) 200-210

[15] Liu Yao, Ning Peng and Reiter M K 2011 False data injection attacks against state estimation in electric power grids[J]. ACM Transactions on Information and System Security 14(1) 21-32

[16] Yu Zonghan and Chin W L 2015 Blind false data injection attack using PCA approximation method in smart grid[J]. IEEE Transactions on Smart Grid 6(3) 1219-1226