Journal of Electronics and Information Science(2017) 2: 122-128
Clausius Scientific Press, Canada

Study of Monitoring False Data Injection Attacks Based on Machine-learning in Electric Systems

Baoyi Wang[a, *], Yadong Zhao[b], Shaomin Zhang[c] , Bihe Li[a]

School of control and Computer Engineering, North China Electric Power University, Baoding 071003, Hebei Province, China
a.email:wangbaoyiqj@126.com, b.email:hdzyd@foxmail.com,
c.email:zhangshaomin@126.com

Keywords: FDIA; Machine Learning; Supervised Learning

Abstract: False data injected by hackers can interfere with power system state estimation and pose a great threat to the safe and reliable operation of modern power systems (FDIA). The traditional bad data detection method can not effectively detect such attacks. In this paper, by extracting relevant power system measurement characteristic value and use the historical data as the sample, using three classical machine learning algorithms (Perceptron, KNN, SVM) of false data injection attack detection, and respectively in IEEE-9, IEEE-57, IEEE-118 simulation platform for test, verify the supervised machine learning algorithm is applied to the validity of false data injection attack detection.

1. Introduction

Electric systems are compound-coupling network systems constituted by physical electric systems and information communication systems. The security and reliability of electric systems impose a great influence on the present society.

As a new type of network attack, false-data injection attacks (FDIA) was first proposed by Yao Liu et al. in 2009 [2]. This type of attack makes full use of the bad-data detecting holes estimated with traditional status and attackers can successfully inject bad data to measurement values and achieve the illegal goals of changing these values and state variables, controlling the running status of electric systems and earning economic interests.

This study mainly focused on the FDIA and monitoring problems in terms of physical respect. We applied sparsely distributed attack models mentioned in literature to simulate FDIA, as well as detecting with two classical machine-learning algorithms.

2. Rationale of false data injection attack

Just like most studies, this research was conducted based on estimation models under direct current, including m pieces of measurement data and n+1 nodes. The estimation model under direct current of electric systems is showed below:

$$T_t = \{(s_i, y_i)\}_{i=1}^{N_{T_t}} \quad z = Hx + \delta \tag{1}$$

The detection of LNR is a classical method in detecting bad data. Assuming that hackers inject false data into measurement data, estimate attack vector a, and induce a state error vector c of state estimation, the residual error can be formulated in equation (2).

$$r_a = \left\| z_a - H\hat{x}_c \right\|_2 = \left\| (z+a) - H(\hat{x}+c) \right\|_2 = \left\| (z - H\hat{x}) + (a - Hc) \right\|_2 \le \left\| z - H\hat{x} \right\|_2 + \left\| a - Hc \right\|_2 = r + \tau_a \tag{2}$$

Where r_a and r represent residual errors with and without false data respectively; τ_a refers to the residual increment caused by false data. Obviously, when a=Hc [2], equation (2) meets Ra= ‖ z-Hx^ ‖ 2=r, namely $\tau_a = 0$, and false data do not influence the residual errors of LNR detection, thereby effectively avoiding the recognition of traditional bad-data detection. Apparently, if attackers are familiar with electric-system network parameters and topological structure and can manipulate specific quantitative measurements, they can build false data which meet $\tau_a = 0$ and manipulate the state-estimation results from electric systems. Meanwhile, $\tau_a = 0$ is not the only way to start an attack. As long as meeting ‖ a-Hc ‖ <r_a-t_a- ‖ z-Hx^ ‖
, state-estimation results can be controlled.

3. Detection by applying machine-learning method

In the given sets of samples $S = \{s_i\}_{i=1}^{M}$ and labels $Y = \{y_i\}_{i=1}^{M}$, $(s_i, y_i) \in S \times Y$ is a Poisson distribution of independent identically distribution. A hypothesis function $f : S \to y$ established, and the relationship between these two parameters is discovered [8]. Attack detecting problem can be defined as a binomial classification problem, in which,

$$y_i = \begin{cases} 1, a_i \neq 0 \\ -1, a_i = 0 \end{cases} \tag{3}$$

$y_i = 1$ Means that the i^{th} measurement value is a false datum, and $y_i = -1$ means that the datum is normal.

3.1. Perceptron method for false data injection attack detection

If providing sample s_i, a Perceptron is adopted by the classification function $f(s_i) = sign(w \bullet s_i)$, in which, $w \in R^{N_i}$ is a weight vector and $sign(w \bullet s_i)$ is defined as follows [9]:

$$sign(w \bullet s_i) = \begin{cases} -1, w \bullet s_i < 0 \\ -1, \end{cases} \tag{4}$$

During the training process, weights are adjusted by every iteration $t = 1, 2..., T$.

$$w(t+1)=w(t)+\Delta w \tag{5}$$

In the equation $\Delta w=\gamma(y_i-f(s_i))s_i$, γ means learning rate. And this algorithm conducts constant iteration until satisfying the stop condition, such as reaching a certain step of a function or a failure threshold. In the verification stage, new samples are verified through function $f(s_i')=sign(w(T)\bullet s_i')$.

3.2. K-NN method for false data injection detection

This algorithm conducts classification among the closest k pieces of sampled values in sample space through sample s_i'[9]. The observed measurement value s_i is treated as a eigenvector. The k pieces of samples $N(s_i'=\{s_{i(1)},s_{i(2)},\ldots,s_{i(k)}\})$ are attained by calculating the distances between samples [10], and $i(1),i(2),\ldots,i(M)$ are defined as follows:

$$\left\|s_i'-s_{i(1)}\right\|_2\le\left\|s_i'-s_{i(2)}\right\|_2\le\ldots\le\left\|s_i'-s_{i(M)}\right\|_2 \tag{6}$$

By calculating the categories of k pieces of the most similar samples, these samples can be classified.

3.3. SVM method for false data injection attack detection

In terms of binomial classification problems, the category of training set $T_i=\{(x_i,y_i)\}_{i=1}^{N_i}$ is yi$\in\{0,1\}$. The separate hyperplane of linear SVM can be achieved by learning [13]: $w\bullet x+b=0$ And the pertinent classification-decision function is:

$$f(x)=sign(w\bullet x+b) \tag{7}$$

Apparently, if the margin is larger, the reliability of classification would be higher (the distance from hyperplane represents the reliability of classification, and the farther of the distance, the more reliable of classification validity). The following function can be easily attained by calculating:

$$margin=\frac{2}{\|w\|} \tag{8}$$

SVM is determined by important training samples (support vectors). Therefore, SVM can be described as the optimization problems of linear classification to amply $\frac{2}{\|w\|}$ to the maximum (equals to minimizing $\frac{1}{2}\|w\|^2$ to the minimum) when all the samples are classified correctly.

$$\min_{w,b}\ \frac{1}{2}|w|^2 \tag{9}$$

$$s.t.\quad y_i(w\bullet x_i+b)-1\ge 0 \tag{10}$$

Introducing Lagrange multiplier ($\alpha_i\ge 0,i=1,2,\ldots,N$) in every inequality constraint to build Lagrange function:

$$L(w,b,a)=\frac{1}{2}|w|^2-\sum_{i=1}^{N}\alpha_i[y_i(w\bullet x_i+b)-1] \tag{11}$$

According to Lagrange allelism, the original problems equal to optimization problems:

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \bullet x_j) - \sum_{i=1}^{N}\alpha_i \tag{12}$$

$$s.t. \quad \sum_{i=1}^{N}\alpha_i y_i = 0 \qquad \alpha_i \geq 0, i = 1,2,...,N \tag{13}$$

In terms of linear problems, linear SVM is not qualified any more, while nonlinear SVM is required. The method of solving nonlinear classification problems is to realize linear separability by spatial transformation (generally means the mapping from low dimension space to high dimension space $x \rightarrow \phi(x)$). The examples in below figures transform the elliptical separate hyperplane of the left figure into the lines in these figures through spatial transformation.

There are inner products of sample points in the objective functions of SVM equivalent dual problems, thereby becoming $\emptyset(x_i) \cdot \emptyset(x_j)$ after the spatial transformation. Because of the increase in dimensions, the calculating costs of inner products increases either, which shows the usability of a kernel function that can transform the mapped inner products in higher dimensional space into a function $(k(x,z) = \emptyset(x) \cdot \emptyset(z))$ in lower dimensional space. Substituting this function into the generalized objective function (7) of SVM learning algorithm, the optimization problems of nonlinear SVM can be attained:

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{N}\alpha_i \tag{14}$$

$$s.t. \quad \sum_{i=1}^{N}\alpha_i y_i = 0 \quad 0 < \alpha_i < C, \quad i = 1,2,...,N \tag{15}$$

4. Simulation analysis

This study adopted IEEE-9, IEE-57 and IEEE-118 bus testing system pairs and conducted simulation analyses to the above three methods respectively. By taking the electric power data published by the Bonneville Electric Power Administration in America as references [12] and assuming that the fluctuating time interval of the total output active power from an electric generator was 5 minutes, the electric power data with a time limit of 5a could be obtained.

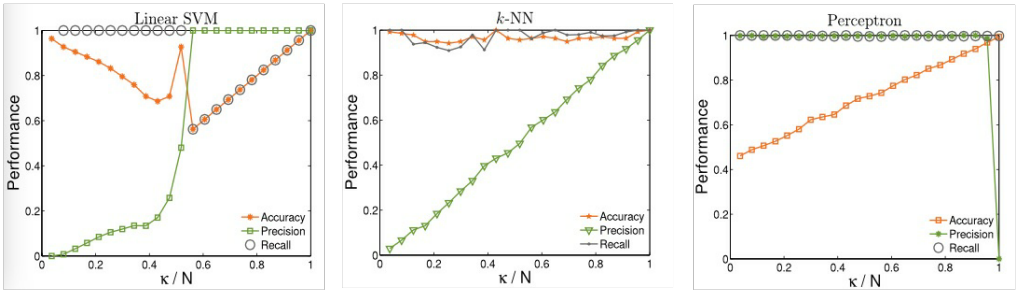

Figure 1 Results for the IEEE 57-bus system. (a) k-NN.
(b) SVM with linear kernel. (c) Perceptron.

We simulated every algorithm on different nodes and observed its results. The result of the three algorithms above simulated on IEEE-57 node is showed in figure 1. It can be suggested from the figure that the precision ratios of perceptron are relatively higher and are not influenced by k/N value. In the algorithm of SVM, it is obvious that with the change of k/N value, the accuracy rates and precision ratios has experienced a significant change. In terms of KNN algorithm, both accuracy rates and precision ratios constantly maintain in a higher level and show no significant fluctuations.

We evaluated the performance of every algorithm according to the accuracy rates and recall rates of false data and normal data and used Class-1 and Class-2 to express the evaluation results.

$$Class-1:\Pr ec-1=\frac{tp}{tp+fp} \tag{16}$$

$$Class-2:\Pr ec-2=\frac{tn}{tn+fn} \tag{17}$$

Where tp refers to that the data are judged as false data by false-data judgment; fp represents that data are judged as false data by normal-data judgment; tn means that data are judged as normal data by normal-data judgment; fn is that data are judged as normal data by false-data judgment.

Figure 2 shows that the precision ratios of perceptron to false-data increase with the increase in k/N, while the precision ratios of both false data and normal data do not change significantly with the variation of k/N and recall rates do not increase with the increase in k/N.

We can see that with the increase in k/N, Class-1 increased gradually, while Class-2 decreased. Thus, k-NN algorithm is sensitive to category homogenization and data sparsity. Moreover, since k-NN algorithm is based on the adjacent samples in Euclidean space and the k/N norms of its attacked measurement values increase with the increase in k/N, decision boundary leans to Class-1.

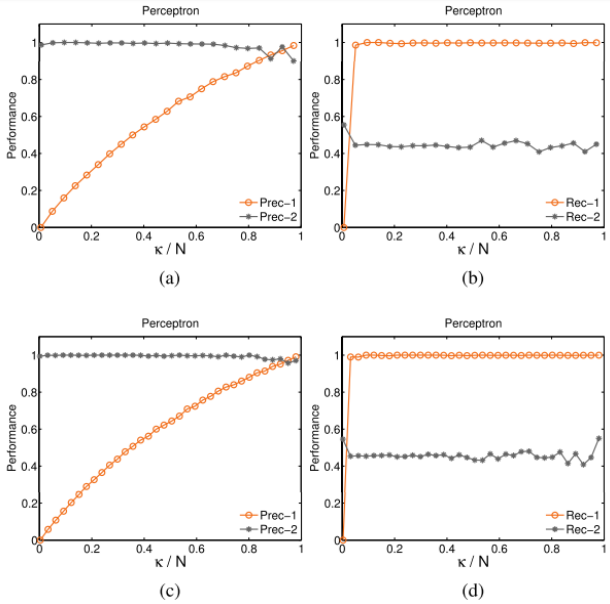

Figure 2 Performance analysis of the Perceptron. (a) Results for the IEEE 57-bus. (b) Results for the IEEE 57-bus.(c) Results for the IEEE 118-bus. (d) Results for the IEEE 118-bus.

5. Conclusion

In the supervised binary classification problems, the attacked and safe measurements are marked as two independent categories. In the experiment, we have observed that machine-learning algorithm shows better performance and can detect FDIA more effectively. Meanwhile, KNN is more sensitive to the size of system than other algorithms. In large-scale systems, the performance of SVM is better than other algorithms. And in the performance test of SVM, we also have observed that phase change κ is the minimum measurement amount required to change when hackers use it to start an attack successfully. Besides, the bigger value of κ does not always means to have a bigger influence on the system. For example, if attack vector a is the smallest among all element values, the influence of a would be extremely limited.

We have observed two challenges in detection problems of SVM when suffering smart power grid attacks. The first is that the performance of SVM is influenced by selection of kernel types. For instance, we have observed that linearity and Gaussian have similar performance in IEEE 9-bus system. However, in terms of IEEE 57-bus system, Gaussian kernel SVM is better than linear SVM. In addition, the values of the phase transformation points of the performance in Gaussian kernel SVM are equal to the theoretical calculating values, which mean that the eigenvector processed by Gaussian kernel function is linearly separable. Secondly, SVM is sensitive to the sparsity of the system. In order to solve this problem, we have applied sparse SVM and kernel machines.

In future studies, we plan to introduce this supervised learning algorithm and non-supervised one into the monitoring of FDIA, and contrast with supervised learning, so that to find out a machine-learning algorithm which is the most suitable one to conduct detection and apply it to false-data detection.

References

[1] Zhao J.H., Liang G.Q., Wen F.S., Dong C.Y. Enlightenment from Ukraine event: Precaution of false-data injection attack to power grids [J]. Automation of Electric Power Systems, 2016,07:149-151.

[2] Liu Y,Ning P,Reiter M K.False data injection attacks against state estimation in electric power grids[J] . ACM Transactions on Information and System

[3] C. Rudin et al., “Machine learning for the New York City power grid,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 2, pp. 328–345, Feb. 2012.

[4] R. N. Anderson, A. Boulanger, W. B. Powell, and W. Scott, “Adaptive stochastic control for the smart grid,” Proc. IEEE, vol. 99, no. 6, pp. 1098–1115, Jun. 2011.

[5] Z. M. Fadlullah, M. M. Fouda, N. Kato, X. Shen, and Y. Nozaki, “An early warning system against malicious activities for smart grid communications,” IEEE Netw., vol. 25, no. 5, pp. 50–55, Sep./Oct. 2011.

[6] Y. Zhang, L. Wang, W. Sun, R. C. Green, and M. Alam, “Distributed intrusion detection system in a multi-layer network architecture of smart

[7] Wang S,Ren W.Stealthy false data injection attacks against state estimation in power systems: Switch ing network to pologies[C]//2014 American Control Conference.Portland,OR:IEEE,2014:1572-1577.

[8] O. Bousquet, S. Boucheron, and G. Lugosi, “Introduction to statis- tical learning theory,” in Advanced Lectures on Machine Learning, O. Bousquet, U. von Luxburg, and G. Rätsch, Eds. Berlin, Germany: Springer-Verlag, 2004.

[9] S. Kulkarni and G. Harman, An Elementary Introduction to Statistical Learning Theory. Hoboken, NJ, USA: Wiley, 2011.

[10] Q. Wang, S.R.Kulkarni, and S. Verdú, “Divergence estimation for mul- tidimensional densities via k-nearest-neighbor distances,” IEEE Trans. Inf. Theory, vol. 55, no. 5, pp. 2392–2405, May 2009.

[11] S. Theodoridis and K. Koutroumbas, Pattern Recognition. Orlando, FL, USA: Academic, 2006.

[12] R. D. Duda, P. E. Hart, and D. G. Stork, Pattern Classification. New York, NY, USA: Wiley