Research on Attack Detection Method of Photovoltaic Grid-connected Interface Device Based on Gradient Boosting Decision Tree

Hao Yao\textsuperscript{1,2}, Shengguo He\textsuperscript{3}, Rui Chen\textsuperscript{3}, Yuansheng Chen\textsuperscript{3}, Yucheng Zheng\textsuperscript{3}, Kai Ding\textsuperscript{3}, Xiaofan Zhu\textsuperscript{3}, Guoqi Zhou\textsuperscript{3}, Zhiwei Huang\textsuperscript{3}, Shengyuan Xiao\textsuperscript{3}, Hongyan He\textsuperscript{3}, Junjun Pan\textsuperscript{3}, Wei Xi\textsuperscript{1,2}, Yang Yu\textsuperscript{1,2} and Baifeng Ning\textsuperscript{4}

\textsuperscript{1}Electric Power Research Institute, China Southern Power Grid. Guangzhou 51000, China
\textsuperscript{2}Digital Grid Research Institute, China Southern Power Grid. Guangzhou 51000, China
\textsuperscript{3}CYG SUNRI CO., LTD. Shenzhen 518000, China
\textsuperscript{4}Shenzhen Power Supply Bureau Co., Ltd, China Southern Power Grid. Guangzhou 51000, China

Corresponding author e-mail: hsguo@sznari.com

Abstract. Photovoltaic grid-connected interface devices are an important class of smart devices in microgrids. The authenticity and reliability of the data they acquire, as well as the safety and stability of operation, are related to the safe and reliable operation of the entire microgrid system. However, in the context of microgrid intelligence and informatization, information/network attacks will become the norm, making network-dependent information interaction methods subject to various security risks. The photovoltaic grid-connected interface device involves an open operating environment and is extremely vulnerable to network attacks. The attack information will occupy the space or resources of the photovoltaic grid-connected interface device, making the photovoltaic grid-connected interface device unable to respond to other important requests or instructions in a timely manner, and in severe cases, will cause the device to be paralyzed and affect the normal system operation. Aiming at the above problems, this paper presents an attack detection method based on the gradient-upgraded decision tree model, and gives a detailed design process of attack detection model of the photovoltaic grid-connected interface device. That is, the important data flow in the photovoltaic grid-connected interface device is used as the input of the gradient-upgraded decision tree model, and then the gradient-upgraded decision tree model detect or classify flows, finally, intercept the data flow with attack behavior and give a warning prompt, and forward data without attack behaviors normally.

1. Introduction
In order to meet the diverse power supply needs of users at this stage, improve the reliability of power consumption, and strengthen the comprehensive utilization of renewable energy sources such as wind energy and solar energy, the concept of microgrid was proposed. Microgrid integrates energy storage devices, power electronic devices, related loads, and some monitoring and protection devices. It can run in island mode, independently supply power to the load, and can also run in grid-connected mode to achieve bidirectional energy flow [1-4].
With the large-scale development and utilization of new energy in domestic and foreign markets, photovoltaic power generation technology [5-7] has also become increasingly mature. And in recent years, the state's support policy for the photovoltaic industry has been continuously introduced, so that photovoltaic power generation technology continues to be improved. On the one hand, photovoltaic power generation has a small scale, a large number, and it is unstable and intermittent depending on conditions such as environmental climate. Therefore, when a large number of photovoltaic power systems are connected to the power grid, it is necessary to obtain timely and accurate information about their operating status. The status data is uploaded to the superior dispatch center via the grid-connected interface device. On the other hand, with the intelligence and informationization of the power grid, a variety of network attack methods are emerging, and photovoltaic grid-connected interface devices involve an open operating environment and are vulnerable to various attacks. These attacks will endanger Information confidentiality, integrity and availability [8-11]. It can be seen from many cases of major power outages caused by cyber attacks or cyber information security incidents in the world and cyber attacks that interfere with the normal operation of the power grid in recent years, and the resulting power outages have been on the rise in recent years, so it is especially important to improve the active attack detection and attack immunity capabilities of microgrids [12-14].

Photovoltaic grid-connected interface devices are an important type of intelligent equipment in microgrids, which not only includes protection, measurement, automatic control, power quality monitoring and other functions, but also an important bridge to communicate with the lower and upper layers. By measuring the operating status information of each photovoltaic power generation unit, on the one hand, it is applied as an internal power quality monitoring system of the device, and on the other hand, it needs to be uploaded to the upper-level control center or dispatch center via the communication module. Therefore, the authenticity and reliability of the data obtained by the photovoltaic grid-connected interface device, as well as the safety and stability of its operation, are related to the safe and reliable operation of the entire microgrid system. In the context of intelligent and informatized power grids, there are various security risks associated with network-based information interaction. However, the photovoltaic grid-connected interface device involves an open operating environment, which allows an attacker to parse the communication protocol within it, or implement eavesdropping attacks, Dos attacks, and tampering with sensitive data [15-17]. As a result, the photovoltaic grid-connected interface device cannot obtain real and effective data; An attacker can also gradually invade the upper-layer control system by using the grid-connected interface device as a springboard.

Early attack detection was mainly based on methods of expert systems, pattern matching, and statistical analysis [18-20]. The attack detection method based on pattern matching has simple algorithm and high accuracy, but it cannot detect unknown attacks; the expert system-based attack detection method depends on the experience of experts and the completeness of the attack library; attack detection based on statistical analysis depends on the setting of the threshold, but also gives the intruder a chance to slowly adapt to the threshold. However, with the development of high-speed networks and advances in attack methods, traditional methods have the disadvantages of high false positive rates, failure to detect unknown attack behaviors, and occupying too many resources. In addition, most of the existing attack detection methods are based on theoretical and model research, and few have considered the possibility of applying them to specific and complex power system operating conditions, such as photovoltaic grid-connected interface devices. In this paper, a machine learning algorithm-based gradient boost decision tree [21-24] is applied to photovoltaic grid-connected interface devices, and a specific attack detection model is designed. This model can realize the active control of the photovoltaic grid-connected interface devices actively and accurately, thereby ensuring the safe and stable operation of the microgrid central controller in the presence of network attacks.

2. Network Attack Analysis of Photovoltaic Grid-connected Interface Devices

Figure 1 shows the structure of a photovoltaic power grid-connected system. The photovoltaic grid-connected interface device collects the voltage and current electrical quantities of the common connection point, and receives information from each inverter controller and load, receives environmental meteorological data and other smart device data from environmental monitors, and
communicates with the photovoltaic monitoring system in the station and the remote dispatching master station. The status of traditional photovoltaic grid-connected interface devices and the operating status of inverters and other smart devices in the system are uniformly monitored through the photovoltaic monitoring system on the station.

However, traditional photovoltaic grid-connected interface devices may have security risks. In general, one or more security vulnerabilities will give an attacker a chance to exploit. The attacker can take this vulnerability as a breakthrough to occupy the network resources of the attacked object and disturb the normal communication by means of creating a large number of useless data or repeatedly sending requests. In traditional photovoltaic grid-connected interface devices, hackers can easily launch various attacks on them based on known security vulnerabilities.

For example, attackers can create a large number of useless data, resulting in network congestion in the grid-connected interface device area, making it unable to communicate with the upper and lower levels normally; The attacker can use the transmission protocol of the grid-connected interface device to deal with the defects of repeated connection, repeatedly send out aggressive repeated connection requests, so that the grid-connected interface device can not deal with other normal requests in time; The attacker can inject Trojan into the grid connection interface device and use it as a springboard to gradually invade the upper system; The attacker can also repeatedly send abnormal attack data according to the transmission protocol defects of the grid-connected interface device, such as tampering with the power consumption data to steal the power charge, real-time price manipulation, tampering with the weather data to cause the environmental monitoring system misjudgment, and further cause the upper controller or the dispatching center to allocate a large number of system resources wrongly, which directly affects the safe and stable operation of the power grid.

**Figure 1. Structure of photovoltaic grid-connected system**

The diagram shows the structure of a photovoltaic grid-connected system. The system consists of photovoltaic arrays, inverters, communication lines, power lines, information collection points, and a dispatch data network. The communication interface devices are connected to the dispatch master station, photovoltaic monitoring system, and environmental monitoring instruments, among other components. The system is designed to monitor and control the photovoltaic arrays and inverters, ensuring safe and stable operation.
3. Attack Detection Method of the Photovoltaic Grid-connected Interface Device Based on GBDT

3.1. The Principle of GBDT Attack Detection
GBDT algorithm-based malicious data attack detection principle is shown in figure 2. GBDT is an ensemble learning algorithm, which is a combination of decision tree and gradient boosting. Therefore, its integration method is gradient boosting. It uses Boosting idea and is composed of a series of integrated weak classifiers. Each weak classifier is given a prediction value, which is combined to form the final prediction value according to a certain weight, so as to get a strong classifier. The GBDT algorithm has obvious advantages in tuning time, classification accuracy, and robustness to outliers.

![Detection principle of malicious data attack based on GBDT algorithm](image)

**Figure 2.** Detection principle of malicious data attack based on GBDT algorithm

In the learning process of GBDT model, first of all, a decision tree weak classifier with fewer leaves is used for iterative learning. Each iteration enlarges the previous learning error, so that the error of the current iteration step is smaller than that of the last iteration, and each iteration will retain all the weak classifiers of the previous iteration step and add a weak classifier with smaller error. Finally, a strong classifier model consisting of several weak classifiers is obtained.

When the GBDT algorithm is applied to the attack detection of the photovoltaic grid-connected device, the characteristic data or influence factors that can indicate whether the photovoltaic grid-connected device is attacked are taken as the input of GBDT. The input data is calculated by each layer of GBDT, and finally the attack type of the photovoltaic grid-connected device is output.

3.2. Design of Attack Detection Model
In this paper, an attack detection method based on convolution neural network for photovoltaic grid-connected device is proposed. The structure of the attack detection system based on GBDT algorithm proposed in this paper is shown in figure 3.
The trained GBDT Classification Model

Positive and Negative Samples

Collecting Data Streams

Data Standardization, Normalization, and Feature Extraction Preprocessing

Training GBDT Model

The trained GBDT Classification Model

Attack Detection

Alarm and Log Records

Figure 3. Structure of attack detection system based on GBDT

It mainly includes the following:(1) Obtain the information data flow of photovoltaic grid-connected device;(2) Preprocess the collected data flow;(3) Train the GBDT classification model, and then input the pre-processed data stream to the trained GBDT classification model for real-time detection and classification, and output the classification results;(4) Intercept or forward the data stream according to the classification results; when there are abnormal classes in the classification results, corresponding alarm prompts and log records are generated according to the classification of the data in the abnormal classes; when the classification results are all normal classes, then forward the data stream.

The specific design steps of attack detection model of photovoltaic grid-connected interface device based on GBDT proposed in this paper are as follows:

Step 1: Obtain the important data flow of the photovoltaic grid-connected interface device. The input data stream of the GBDT-based photovoltaic grid-connected interface device attack detection model includes the voltage, current, frequency, active power, reactive power, and power factor of the common connection point; Remote signalling, telemetry, power consumption, and commands for remote opening and closing, remote adjustment, start and stop; Output power of photovoltaic power generation unit, as well as load power, environmental meteorological data (temperature, light intensity);

Step 2: Preprocess the data stream. It mainly includes feature extraction, numericalization and normalization. Feature extraction refers to the extraction of features that best characterize the operating status of a photovoltaic grid-connected interface device, such as the number of requests and instructions sent by a certain smart device to the photovoltaic grid-connected interface device, or the number of wrong data sent by a device to the photovoltaic grid-connected interface device, or the number of data sent by a device to the photovoltaic grid-connection interface device under different communication protocols. The complexity of neural network model training can be reduced and the accuracy of model detection can be improved by extracting the features that can best represent the operation state of photovoltaic grid connection interface device.

Actually obtained data has both numeric type variables and character type variables (such as communication protocol types), so it is necessary to numerically process character type variables.

After the numerical processing of the data, there are large differences in the values between the features, which makes it easy for large features to cover small features, which is not conducive to the speed and accuracy of neural network training. Therefore, it is necessary to normalize the features and map them to the [0,1] interval.
The normalization formula is as follows:

\[ y = \frac{x - \min}{\max - \min} \]

(1)

Where \( x \) and \( y \) are the feature values before and after normalization, and \( \max \) and \( \min \) are the maximum and minimum values of each feature.

Step 3: First, train the GBDT classification model. Before training, a training sample set is established, the sample set includes positive samples and negative samples, the positive samples are normal data streams, and the negative samples are data streams that have been maliciously attacked. The malicious attacks include Dos attacks, unauthorized access attacks, abnormal detection on the interface side, Trojan horse attacks, operation status, forgery or tampering of messages such as running state and weather. The specific training process is:

1. According to the label value \( y_i \) set in the sample set (including positive and negative samples), the weak learner model is initialized:

\[ f_k(x) = 0, k = 1, 2, ..., K \]  

(\( K \) is the number of categories, and \( K = 6 \));

2. Set the number of iterations \( m = 1, 2, ..., M \):

   2.1) Calculate the probability that the sample points belong to each category:

\[ P_k(x) = \frac{\exp(f_k(x))}{\sum_{l=1}^{K} \exp(f_l(x))} \]

(2)

Where \( \exp(f_k(x)) \) represents the index of \( f_k(x) \); \( K \) is the number of categories;

2.2) For each classification type \( k = 1, 2, ..., K \):

   2.2.1) Calculate the residual:

\[ r_{ki} = y_{ki} - P_k(x_i) \]

(3)

Where \( i = 1, 2, ..., n \) is the number of samples; \( y_{ki} \) is the value of class \( K \) corresponding to the \( i \)th sample; \( P_k(x_i) \) is the probability that sample \( x_i \) belongs to class \( k \);

   2.2.2) Retraining to fit a classification tree with probability pseudo residuals \{ \( (x_1, r_{k1}) \), ..., \( (x_n, r_{kn}) \) \};

   2.2.3) Calculate the multiplier:

\[ c_{mkj} = \frac{K - 1}{K} \frac{\sum_{x_i \in R_{mkj}} r_{ki}}{\sum_{x_i \in R_{mkj}} \left| r_{ki} \right| \left(1 - \left|r_{ki}\right|\right)} \]

(4)

Where \( K \) is the number of categories; \( c_{mkj} \) is the leaf node multiplier of the tree generated by \( m \) iterations and \( k \)-classes; \( j = 1, 2, J \) is the number of leaf nodes; \( m = 1, 2, M \) is the number of iterations; \( x_i \) (\( i = 1, 2, ..., n \)) is the input sample; \( r_{ki} \) is the class \( k \) pseudo residual of the \( i \)th sample;

   2.2.4) Update the learner according to equation(5):

\[ f_{k,m}(x) = f_{k,m-1}(x) + \sum_{j=1}^{J} c_{mkj}I(x \in R_{mkj}) \]

(5)

Where \( f_{k,m}(x) \) is the \( m \)-th iteration of the sample \( x \) and the \( k \)-th class learner; \( f_{k,m-1}(x) \) is the \( m \)-1th iteration of the sample \( x \) and the \( k \)th class learner; \( I \) is the set of leaf features; \( R_{mkj} \) is the leaf node region of the tree generated by \( m \)-th iteration and class \( k \);

3) Output the strong classifier:
\[ F_{Mk}(x) = \sum_{m=1}^{M} \sum_{j=1}^{J} c_{mjk} l(x \in R_{mkj}) \]  

(6)

Where \( F_{Mk}(x) \) is the strong classifier obtained by \( M \) iterations of the sample \( x \) and the \( k \) class;

The final \( F_{Mk}(x) \) is used to obtain the corresponding probability \( P_{Mk}(x) \) of class \( k \):

\[ P_{Mk}(x) = \exp(F_{Mk}(x)) \sum_{l=1}^{K} \exp(F_{Ml}(x)) \]  

(7)

Convert probability to category:

\[ \hat{k}(x) = \arg \min_{1 \leq k \leq K} \sum_{k'=1}^{K} [c(k, k')P_{Mk'}(x)] \]  

(8)

Where \( \hat{k}(x) \) is the final output category, \( c(k, k') \) is the joint cost when the real value is \( k' \) and the prediction is class \( k \), that is, the category with the greatest probability is the predicted category.

After the GBDT classification model is trained, the preprocessed data stream is classified using the trained GBDT model, and the classification results are output. The classification includes normal classes and malicious attacks. The malicious attacks include Dos attacks. Authorized access attacks, abnormal detection on the interface side, Trojan horse attacks, operation status, forgery or tampering of messages such as running state and weather.

Step 4: Intercept or forward the data flow according to the classification results. When there is an attack class in the classification result, the corresponding alarm prompt will be sent according to the classification of the data in the attack class, and a log record will be generated to save and the data flow will be intercepted; When the classification results are all normal, the data flow will be forwarded.

### 3.3. Application of Attack Detection Model

In this paper, the GBDT classification model is implemented by adding an auxiliary CPU. The auxiliary CPU is connected to the main CPU to complete the detection of the data flow of the photovoltaic grid-connected interface device. Figure 4 shows the hardware structure of the photovoltaic grid-connected interface device.

![Hardware structure of photovoltaic grid-connected interface device](image)

**Figure 4.** Hardware structure of photovoltaic grid-connected interface device
It mainly includes:

1. Main control module: used to connect and communicate with the upper layer (the photovoltaic monitoring system and the control master station in the station) and the lower layer (inverter controllers, environmental meteorological monitoring devices and other intelligent devices) via the communication module. At the same time, it receives the information data flow, classification results, and alarm prompts sent by the attack detection module and then sends them to the display module for display; and sends the attack alarm information to the upper layer (on-site photovoltaic monitoring system and control master station) through the communication module;

2. Display module: it is used to display the information for display and alarm information from the photovoltaic grid-connected interface device. The display module can be a display or an indicator, or a combination of both to display and provide more information to achieve better human-computer interaction experience; The alarm prompt is to display the type of malicious attack through a display or an indicator, and when the indicator is used, different colored light sources correspond to different types of malicious attacks;

3. Storage module: it is used to store alarm prompts, log records, control programs, voltage, current, and other electrical parameter information of public connection points, state information of public connection point switches, load switching switches, and circuit breaker switches of photovoltaic power generation systems, and User information, etc.; The log records include attack time, attack duration, attack mode, type of transmission protocol corresponding to the attack, incorrect data segmentation, start and end address information of the incorrect data (i.e., source and target device address information), etc.;

4. Output module: used to output command signals for controlling public connection point switches, load switching switches, circuit breaker switches in photovoltaic power generation systems;

5. Input module: used to receive public connection point switches, load switching in microgrid;

6. AC acquisition module: used to collect analog quantities such as voltage and current at the common connection point, as well as the micro source and load data in the microgrid, and then the collected analog signals are converted into digital signals that can directly participate in the calculation;

7. Communication module: it is used to exchange data with lower-level controllers such as inverter controllers and environmental meteorological monitoring devices in photovoltaic power generation systems, and upper-level controllers such as photovoltaic monitoring systems and control master stations. These data include the instruction information of the photovoltaic monitoring system and the master station in the upper station, remote signaling, telemetry, power consumption and other data, remote opening and closing, remote adjustment, start and stop commands, the output power, load power, Environmental meteorology (temperature, light intensity), etc. In addition, the communication module also sends the output power, load power, environmental weather (temperature, light intensity) of the photovoltaic power generation unit to the attack detection module;

8. Power module: used to provide working power for each module;

9. Attack detection module: it is used to preprocess the information data stream composed of the input module, AC acquisition module, and information data sent from the main control module and communication module, and then use the GBDT model to classify and output classification results. The classification includes normal classes and malicious attack classes. When an attack class exists in the classification result, then a corresponding alarm prompt is issued to the main controller according to the classification of the data in the malicious attack class, and a log record is generated and the data will be intercepted at the same time; When the classification results are all normal, the information data stream is sent to the main control module, and the main control module sends the information data stream to the upper layer through the communication module.

A typical case of the photovoltaic grid-connected interface device with GBDT algorithm attack identification function is shown in figure 5. If the attacker makes a malicious data attack on the illumination intensity information of the photovoltaic power station, which makes the data abnormal to the normal value, the attack detection module in the photovoltaic grid-connected interface device will identify the abnormal behavior from the communication module to the illumination intensity data of the main CPU, so as to output the malicious data attack detection results.
Figure 5. Typical case of photovoltaic grid-connected interface device with GBDT algorithm attack recognition function

4. Conclusions

There are security loopholes in the traditional photovoltaic grid-connected interface device: (1) The status monitoring module of the traditional photovoltaic grid-connected interface device is limited to monitoring the grid connection switch tripping, voltage transformer and current transformer disconnection, power and current overrun, and does not consider monitoring the malicious data attack;(2) The traditional photovoltaic grid-connected interface device lacks an active attack identification module and cannot respond to the rapidity of information intrusion and the delay of communication. Therefore, when the abnormal state information of the photovoltaic grid-connected interface device is uploaded to the photovoltaic monitoring system in the station, the attacker may have launched a deeper attack on the photovoltaic grid-connected interface device or has gradually invaded the upper-layer system.

Therefore, aiming at the problem that photovoltaic grid-connected interface devices in open operating environments are highly vulnerable to network attacks, a method of attack detection based on the gradient-upgraded decision tree model is proposed in this paper, and gives a detailed design process of attack detection model of the photovoltaic grid-connected interface device. That is, the important data flow in the photovoltaic grid-connected interface device is used as the input of the gradient-upgraded decision tree model, and then the gradient-upgraded decision tree model detect or classify flows, finally, intercept the data flow with attack behavior and give a warning prompt, and forward data without attack behaviors normally.

5. Acknowledgments

This work was by Supported by the National Key R&D Program of China(2018YFB0904900, 2018YFB0904903).

6. References

[1] Shafiee Q, Nasirian V,Vasquez J C, Guerrero JM and Davoudi A 2018 A multi-functional fully distributed control framework for AC microgrids IEEE Transactions on Smart Grid 9 3247
[2] Lopes J A P, Moreira C L and Madureira A G 2006 Defining control strategies for microgrids islanded operation IEEE Transactions on Power Systems 21 916
[3] Haider S, Li G and Wang K 2018 A dual control strategy for power sharing improvement in islanded mode of AC microgrid Protection Control Mod. Power Syst. 3 10
[4] Gupta A, Doolla S and Chatterjee K 2018 Hybrid ac-dc microgrid: Systematic evaluation of control strategies IEEE Transactions on Smart Grid 9 3830
[5] Dutta S and Chatterjee K 2018 A buck/boost based grid connected PV inverter maximizing power yield from two PV arrays in mismatched environmental conditions IEEE Transactions on Industrial Electronics 65 5561
[6] Roman E, Alonso R and Ibáñez P 2006 Intelligent PV module for grid connected PV system IEEE Transactions on Industrial Electronics 53 1066
[7] Campanhol L B G, Silva S A O, Oliveira A A and Bacon V D 2019 Power flow and stability analyses of a multifunctional distributed generation system integrating a photovoltaic system with unified power quality conditioner IEEE Transactions on Power Electronics 34 6241
[8] Kim S H, Wang Q H and Ullrich J B 2012 A comparative study of cyberattacks Communications of the ACM 55 66
[9] Madan B B and Banik M 2014 Attack tolerant architecture for big data file systems ACM SIGMETRICS Performance Evaluation Review 41 65
[10] Al-Saleh M I, AbuHjeela F M and Al-Sharif Z A 2015 Investi-gating the detection capabilities of antiviruses under concurrent attacks International Journal of Information Security 14 387
[11] Gamage T T, Roth T P, McMillin B M and Crow M L 2013 Mitigating event confidentiality violations in smart grids: An information flow security-based approach IEEE Transactions on Smart Grid 4 1227
[12] Stamp J, McIntyre A and Richardson B 2009 Reliability impacts from cyber attack on electric power systems IEEE Power Systems Conference and Exposition(USA:Seattle, WA) p 1
[13] Jin D, Li Z, Hannon C, Chen C, Wang J, Shahidehpour M and Lee C W 2017 Toward a cyber resilient and secure microgrid using software-defined networking IEEE Transactions on Smart Grid 8 2494
[14] Nourian A, Madnick S 2018 A systems theoretic approach to the security threats in cyber physical systems applied to Stuxnet IEEE Trans. Dependable Secure Comput. 15 2
[15] Fang H, Xu L, Zou Y, Wang X and Choo K -K R 2018 Three-stage stackelberg game for defending against full-duplex active eavesdropping attacks in cooperative communication IEEE Transactions on Vehicular Technology 67 10788
[16] Zhang H, Cheng P, Shi L and Chen J 2016 Optimal DoS attack scheduling in wireless networked control system IEEE Trans. Control Syst. Technol. 24 843
[17] Sedghi H and Jonckheere E 2013 Statistical structure learning of smart grid for detection of false data injection IEEE Power & Energy Society General Meeting (Canada:Vancouver, BC,) p 1
[18] Denning D E 1987 An intrusion-detection mode IEEE Transactions on Software Engineering 13 222
[19] Chen Z, Zhang Y, Chen Z and Delis A 2009 A digest and pattern matching-based intrusion detection engine The Computer Journal 52 699
[20] Ye N, Emran S M, Chen Q and Vilbert S 2002 Multivariate statistical analysis of audit trails for host-based intrusion detection IEEE Trans. Comput. 51 810
[21] Friedman J H 2002 Stochastic gradient boosting Computational Statistics & Data Analysis 38 367
[22] Schonlau M 2005 Boosted regression (boosting): An introductory tutorial and a Stata Plugin The Stata Journal: Promoting communications on statistics and Stata 5 330
[23] Chawla N V, Bowyer K W, Hall L O and Kegelmeyer W P 2011 SMOTE: Synthetic Minority Over-sampling Technique Journal of Artificial Intelligence Research 16 321
[24] Friedman J H 2001 Greedy Function Approximation: A Gradient Boosting Machine The Annals of Statistics 29 1189