A Combinatorial Optimization Analysis Method for Detecting Malicious Industrial Internet Attack Behaviors

KEJING ZHAO

Henan University of Science and Technology, China

ZHIYONG ZHANG*

Henan University of Science and Technology, China

KIM-KWANG RAYMOND CHOO

University of Texas at San Antonio, USA

ZHONGYA ZHANG

Henan University of Science and Technology, China

TIANTIAN ZHANG

Henan University of Science and Technology, China

Industrial Internet plays an important role in key critical infrastructure sectors and is the target of different security threats and risks. There are limitations in many existing attack detection approaches, such as function redundancy, overfitting and low efficiency. A combinatorial optimization method Lagrange multiplier is designed to optimize the underlying feature screening algorithm. The optimized feature combination is fused with random forest and XG-Boost selected features to improve the accuracy and efficiency of attack feature analysis. Using both the UNSW-NB15 and Natural gas pipeline datasets, we evaluate the performance of the proposed method. It is observed that the influence degrees of the different features associated with the attack behavior can result in the binary classification attack detection increases to 0.93, and the attack detection time reduces by 6.96 times. The overall accuracy of multi-classification attack detection is also observed to improve by 0.11. We also observe that nine key features of attack behavior analysis are essential to the analysis and detection of general attacks targeting the system, and by focusing on these features one could potentially improve the effectiveness and efficiency of real-time critical industrial system security. In this paper, CICDDoS2019 dataset and CICIDS2018 dataset are used to prove the generalization. The experimental results show that the proposed method has good generalization and can be extended to the same type of industrial anomaly data sets.

CCS CONCEPTS • Security and privacy → *Industrial Internet security***;** I**ntrusion detection systems; • Computing methodologies → Feature selection**;

Additional Keywords and Phrases: Industrial Internet, Industrial situational security, Attack behavior, Feature analysis, Combinatorial optimization.

1 INTRODUCTION

The role of Industrial Internet is increasingly important with the global industrial digitalization, networking and intelligent process, and this is partly evident in the global trend (e.g., German Industrial 4.0, Japanese Society 5.0, USA Industrial Internet and China Intelligent Manufacturing 2025 initiatives). Generally, Industrial Internet-based systems integrate both Information Technology (IT) and Operational Technology (OT)–also referred to as Industrial Internet of Things (IIoT), Industrial Cyber-Physical Systems, or Industrial Human-centric Cyber Physical (Production) Systems in the

ACM 2378-962X/2023/12-ART
http://dx.doi.org/10.1145/3637554

literature [1]. The Industrial Internet covers the technical part of IIoT, the industrial Internet is to achieve the comprehensive interconnection of human, machine and things, and the pursuit of business digitization; The IIoT emphasizes the connection between things and things, and the pursuit of production automation. Industrial Human-centric Cyber Physical (Production) Systems. Compared with Industrial CPS, Industrial Human-centric Cyber Physical (Production) Systems emphasizes the role of human more, highlighting the synergistic effect of human in production automation, and forming a Human-centric complete closed loop. [2] proposes a method to grade the impact of the industrial attack, complete the feasibility assessment of the attack, and then analyze the overall risk profile. [3] studies the security synthesis of cyber–physical systems subject to stealthy attacks via zonotopic set theory such systems are increasingly deployed in sectors such as advanced manufacturing, energy, electricity, and medical treatment.

The cyber attack refers to any type of offensive action against a computer information system, infrastructure, computer network, or personal computer equipment. For the Industrial Internet, an action that causes a device or service to lose functionality is considered an attack. There have been numerous attempts in designing security solutions, such as those to analyze and detect suspicious network behavior. For example, principal component analysis (PCA) and linear discriminant analysis (LDA) can be used in feature dimension reduction to conduct preliminary feature processing [4]. [5] proposed a novel iterative bi-level hybrid intelligence model named ManuChain to get rid of unbalance/inconsistency between holistic planning and local execution in individualized manufacturing systems. In recent years, machine learning has also been widely used to facilitate the analysis of Industrial Internet traffic features. The importance of feature processing methods such as decision tree, linear regression, random forest, Lasso regression method or their improved versions [6] is particularly pronounced in delay-sensitive settings.

One ongoing challenge is the large amount of network data in the Industrial Internet setting, which results in a significant increase in the training cost and time. However, Industrial Internet has a high requirement on the detection time of attack categories. Considering the real-time needs of similar communication, the requirement of low delay becomes an urgent challenge to be solved. Although it is possible to explore the influence of different features on the final attack results by adding them one by one [12],[31],[34], this method is time-consuming and ignores the interaction between features. For example, pairs can cause certain features to fail.

Aiming at the shortcomings and limitations of existing methods, in this paper, irrelevant variables are eliminated based on the feature processing method, and the feature combination after optimizing the decision tree is compared with the feature combination added one by one. The effectiveness of the new optimization method is proved through the binary and multi-classification experiments on the attack results.

A summary of our contributions is as follows:

1) A combinatorial optimization method Lagrange multiplier is designed to optimize the feature screening algorithm, and the optimized feature combination is fused with the features selected by random forest and XG-Boost.
2) We observe that nine key features of attack behavior analysis are essential to the analysis and detection of general attacks targeting the system. The nine key features found in this paper play an important role in the detection and

identification performance of industrial Internet attacks. The combination of the nine optimal features can achieve extremely high detection performance and greatly reduce the running time.

3) The experimental results show that the proposed method has good generalization and can be extended to the same type of industrial anomaly data sets.

The rest of this paper is organized as follows. Section 2 discusses the related work, and Section 3 describes the characteristics of attack behavior(s) associated with Industrial Internet-based systems, prior to briefly introducing the principle of Lagrange multiplier optimization decision tree algorithm. Then, Section 4 introduces the evaluation setup. Section 5 discusses our findings. Finally, Section 6 concludes this research.

2 RELATED WORK

Industrial Internet situational security is key in many of our critical infrastructure sectors, and we need to maintain a watchful brief on the fast evolving cyber threats landscape.

2.1 Industrial scenarios attack data imbalance

With the continuous expansion of the scale of industrial systems, the stratification of industrial control systems also represents the differentiation of industrial data generated at different levels. The data of industrial systems are mostly equipment data, instruction data, status data, etc., and the traffic data generated by their attack behavior also presents two characteristics: structured and unstructured. Unstructured data refers to data whose data structure is irregular or incomplete, and it is not convenient to use two-dimensional logical tables in the database. Two-dimensional representations can effectively regularize the data in order to obtain more information in the data The unstructured nature of the data makes it often nonlinear and high-dimensional, making it impossible for the above methods to handle large amounts of such data. At the same time, due to the serious imbalance of network attack data in industrial control system, that is, most of the collected data is the normal operation data of the network layer, and the actual attack data only accounts for a very small part, and the data distribution is uneven. At present, the treatment methods for data imbalance mainly include oversampling, under-sampling and SMOTE. Oversampling refers to expanding a small sample size to improve model performance, but this method also introduces overfitting problems; Under-sampling refers to the number of deleted samples, but also increases the probability of losing important samples; SMOTE increases the probability of class boundaries overlapping and usually generates some useless samples. In view of this, [22] studies the selection method of event traffic characteristics of Industrial Internet, establishes the propagation evolution model of security events in Industrial Internet, and achieves the traceable mapping of security event propagation. [23] This paper analyzes the current application of artificial intelligence methods in the field of industrial Internet, and focuses on the key technologies of artificial intelligence in industrial Internet security and the existing defects and deficiencies. In order to prevent the impact of potential cyber attacks, [24] proposed the structured classification of key industrial assets in Industry 4.0 and the potential adverse impact of cybersecurity vulnerabilities on business performance, and explored the loss analysis of data confidentiality, integrity, and availability associated with networked manufacturing machines. Considering that it is difficult for existing attack detection methods to simultaneously take known attacks and unknown

attacks in the industrial Internet of Things into account, [25] a deep learning-based intrusion detection paradigm is proposed, which adopts mixed rule-based feature selection to train and verify information captured from TCP/IP packets. The scheme was tested using NSL-KDD and UNSW-NB15 datasets.[26] An integrated framework of smart grid intrusion detection system (IDS) is proposed based on the combination of feature engineering preprocessing and machine learning classifiers. This method focuses on selecting more important features by using gradient enhanced feature selection method before applying classification algorithm, and the effectiveness of this method is proved through experiments.[27] explored the further integration of deep learning and Industrial Internet by using PCA and other methods for feature dimension reduction based on the characteristics of Industrial Internet. Among them, [28] designed a bi-directional long short-term memory network (B-MLSTM) with multi-feature layers. Considering its redundancy and coupling between attack data not only increases the training time and overfitting of the algorithm, but also reduces the detection accuracy.

2.2 Industrial scenario attack detection model

Nowadays, machine learning methods have been widely used in Industrial Internet network attack detection and attack intention recognition, and such systems are often delay-or time-sensitive. [7] used machine learning methods to build industrial anomaly detection models. There have been a great deal of intrusion detection systems (IDS) designed in the literature, for example the authors of [8] respectively proposed network remote sensing, hybrid machine learning IDS architecture, generalized learning system to facilitate network attack behavior detection and attack intention recognition. [11] proposed a method according to the evaluation on the cross-enterprise collaboration. Around the intelligent manufacturing security issues, [12] from the aspect of social situation, information physical system and manufacturing can be combined to establish a new security platform. Meanwhile, for vulnerability detection in network attacks, the authors of [14] explored the characteristics and rules of network vulnerability attacks in industrial situational by studying vulnerability detection in industrial operating systems. [15-16] analyzed the denial-of-service attacks in industrial scenarios and the feature analysis methods and detection recognition in SCADA systems. Hence, the authors of [17] constructed algorithms to improve the sensitivity of detection systems using different machine learning methods to facilitate low-cost decision-making. New privacy perception system and MTD algorithm proposed in the literature, the strategy by gathering the attacker in the same service access to achieve the optimal allocation, supporting service migration decisions on value iteration, obtained the consistent decision.

2.3 Algorithm improvement of attack classification model

From the algorithm level, the analysis of network attack behavior characteristics in Industrial Internet scenarios mainly improves the classification performance through the algorithm training process and the use of multiple integrated strategies. The attack categories in industrial scenarios are complex and diverse, and the industrial scenarios such as communication have strong real-time performance, which puts forward new requirements for the detection time of attack categories. Based on the system sensitivity algorithm, [18] using neural network method to detect new attacks, the overall performance of attack detection is improved. Industrial Internet has a lot of data, network security

data synchronization attacks would greatly increase the cost of training, aiming at this problem, [19] proposes a new method for detecting IIoT synchronous attacks according to feature processing based on feature processing. Aiming at the state estimation problem of smart grid, [20] proposed an attack detection method based on XG-Boost. By training and adjusting the parameters of XG-Boost model, data features were extracted and a high-precision false data injection attack detection model was realized. [21] uses open source data sets and simulators to carry out real equivalent simulations, which also proves the feasibility of the method. The research on the characteristics of Industrial Internet network attack behavior based on the algorithm level mainly improves the classification performance by improving the algorithm training process and adopting a variety of integration strategies. Scholars begin to study the fusion of two or more machine learning methods. For example, decision trees were combined with Naive Bayes [29], multiple machine learning methods were integrated [30], DNA computing was combined with decision trees [31]. [32] introduced a scalable model-based method to reveal generic attacks, and further deduced to identify attacks by solving sparse optimization problems.

In summary, with regard to the feature analysis and attack classification research of cyber attack behavior in industrial scenarios, many scholars still directly use high-complexity and high-dimensional raw data to explore, and the influence of different features on the final attack result is mainly explored by adding features one by one [9],[28],[31], but this method ignores the interaction between features. The feature quantity of industrial data is too large, which increases the computational amount of attack classification algorithm, reduces the detection efficiency, and raises the requirements for experimental equipment, which is very unfavorable to the security analysis in industrial scenarios Since attacks in industrial scenarios mostly involve equipment, network, physical state, etc., attack characteristics are highly interactive, and there may be some intrinsic relationship between features, and the processing and analysis of original attack data and the exploration of efficient attack classification methods are imminent.

3 ANALYSIS OF THE CHARACTERISTICS OF ATTACK BEHAVIOR IN INDUSTRIAL SITUATIONAL SECURITY

In this section, we propose a feature optimization algorithm. The notations used in the following discussion are summarized in Table 1.

Table 1: Notations

Symbol	Meaning	Symbol	Meaning
x_n	The input data	$\Lambda(\cdot)$	The indicator function
y	The output label	k	Number of categories
λ_i	The Lagrange multiplier	v	Number of branch node
φ	The constraint function	w	Branch node weight
f	The target function	Ent	Information entropy
a	Any constant	$Gain$	Information gain

3.1 Lagrange optimization algorithm for attack feature analysis

The attack behavior feature plays a key role in the detection of network attacks. The decision tree model trained on the balanced data set can't adapt to the heterogeneous data types of the non-balanced data set, which leads to the deterioration of the subsequent anomaly detection performance. In the process of feature analysis of industrial Internet network attack behavior, the decision tree classification model has strong processing ability for irrelevant data, so it can well adapt to the feature diversity of attack behavior under the industrial Internet, feature diversity means that the industrial data generated by different layers of the industrial control system often shows greater differences, and the data categories and formats are more diverse. However, due to the imbalance of sample data in most industrial datasets, for features that contain only a small number of samples under the branch, the purity of the decision tree at the branch node has lost its value, so that such a decision tree does not have the ability to generalize. In view of this situation, this paper proposes to use the Lagrange multiplier method to optimize the ensemble classification algorithm based on decision tree. The Lagrange multiplier is one of the combinatorial optimization methods that transforms an optimization problem with *n* variables and *k* constraints into an extreme problem of an equation system with *n+k* variables whose variables are not constrained in any way. The Lagrange multiplier method has good performance in convex optimization problems, superior convergence performance, and can converge to the global minimum point at a fast speed, which is suitable for the special requirements of attack classification time in industrial scenarios. Considering that the algorithm needs to find the partial derivative of all variables one by one, there is a large computational complexity, but the decision tree algorithm selected in this paper has a small computational complexity (specifically the logarithm of data points for training the decision tree), which can effectively alleviate the computational complexity problem of the Lagrange multiplier method.

The number of network attack behavior characteristics in industrial situational security is n, which is characteristics set to $x_1, x_2, \cdots x_i, \cdots, x_n$ respectively, take the set $X = (x_1, x_2, \cdots, x_n)$ to construct the Lagrange function as follows:

$$L(X, \lambda_i) = f(X) + \sum_{i=1}^{n} \lambda_i \phi_i(X) \quad (1)$$

The algorithm is integrated with decision tree-based learners by optimizing the loss function using the Lagrange multiplier method. Since the division of feature nodes uses information gain in the decision tree, the cross-entropy loss function is selected as the target function f, the output label is $\hat{y}$, and the real label is y, $I = (1, 2, \cdots, n)$, then:

$$f(\hat{y}_i, y_i) = -y_i \log(\hat{y}_i) - (1 - y_i) \log(1 - \hat{y}_i), \quad i \in I. \quad (2)$$

Then the corresponding Lagrange function is:

$$L(\hat{y}_i, y_i) = f(\hat{y}_i, y_i) + \sum_{i=1}^{n} \lambda_i \varphi_i(\hat{y}_i, y_i), \qquad i \in I. \quad (3)$$

Where the constraint function $\varphi_i(\hat{y}_i, y_i)$ is:

$$\phi_i(\hat{y}_i, y_i) = \Lambda(\hat{y}_i = y_i), \qquad i \in I. \quad (4)$$

Where $\Lambda(\cdot)$ is the indicator function and takes the value 0,1 when is true and false. The function form with the constraint is:

$$\begin{cases} \dfrac{\partial f}{\partial x_1}+\sum_{i=1}^{n}\lambda_i\dfrac{\partial \varphi_i}{\partial x_1}=0 \\ \dfrac{\partial f}{\partial x_2}+\sum_{i=1}^{n}\lambda_i\dfrac{\partial \varphi_i}{\partial x_2}=0 \\ \vdots \\ \dfrac{\partial f}{\partial x_n}+\sum_{i=1}^{n}\lambda_i\dfrac{\partial \varphi_i}{\partial x_n}=0 \\ \varphi_1\left(x_1,x_2,\cdots,x_n\right)=0 \\ \varphi_2\left(x_1,x_2,\cdots,x_n\right)=0 \\ \vdots \\ \varphi_n\left(x_1,x_2,\cdots,x_n\right)=0 \end{cases} \quad (5)$$

Let the first partial derivative of $L(\hat{y}_i, y_i)$ with respect to $\hat{y}_i$, y_i, and λ_i be zero, according to the chain rule.

$$\begin{aligned} L'_{\hat{y}_i} &= \frac{\partial}{\partial \hat{y}_i} f(\hat{y}_i, y_i)+\sum_{i=1}^{n}\lambda_i\frac{\partial}{\partial \hat{y}_i}\varphi(\hat{y}_i, y_i) \\ &= -y_i\frac{1}{\hat{y}_i \ln a}-(1-y_i)\frac{-1}{(1-\hat{y}_i)\ln a}+\sum_{i=1}^{n}\lambda_i\varphi'_{\hat{y}_i}(\hat{y}_i, y_i) \\ &= \frac{-y_i}{\hat{y}_i \ln a}+\frac{1-y_i}{(1-\hat{y}_i)\ln a}+\sum_{i=1}^{n}\lambda_i \arg\min(\hat{y}_i - y_i) \\ &= 0 \end{aligned} \quad (6)$$

Where a is any constant, the following equations can be obtained similarly:

$$\begin{cases} L'_{\hat{y}_i} = -y_i(\hat{y}_i \ln a)^{-1}+(1-y_i)[(1-\hat{y}_i)\ln a]^{-1}+\sum_{i=1}^{n}\lambda_i \arg\min(\hat{y}_i - y_i)=0 \\ L'_{y_i} = -\log_a(\hat{y}_i)+\log_a(1-\hat{y}_i)+\sum_{i=1}^{n}\lambda_i \arg\min(\hat{y}_i - y_i)=0 \\ L'_{\lambda_i} = \sum_{i=1}^{n}\arg\min(\hat{y}_i - y_i)=0 \end{cases} \quad (7)$$

Then the extreme point (stagnation point) of $f_i(\hat{y}_i, y_i)$ under the constraint function $\varphi_i(\hat{y}_i, y_i)$ can be obtained according to the above equations. The purity of decision tree node is obtained at this extreme point to avoid the convergence to the local minimum point caused by the unbalance of sample data due to too small sample size.

The decision tree optimized by Lagrange multiplier method can select features according to the importance evaluation of features, and train features more related to the objective function based on the algorithm, which can not only shorten the calculation time, improve the operation efficiency, but also improve the prediction accuracy and overall performance of the model.

Consistent with the above optimization scheme, taking the Industrial Internet network attack feature set X, the information gain was used to assess the importance of the features, the information entropy at the n node of the m tree is

$$Ent(p) = -\sum_{k=1}^{K} p_{nk} \log_2 p_{nk} \quad (8)$$

Where k represents K categories and p_{nk} represents the proportion of category k in node n.

For feature x_i, there are V possible values, which can be expressed as $\{x_{i1}, x_{i2}, \cdots, x_{iV}\}$. The sample set is divided by x_i, and X^V means that the v branch node contains all the samples in X whose value is v in attribute x_i.Considering that different branch nodes contain different samples, weight $w_n = |X^v| / |X|$ is assigned to branch nodes, that is, branch nodes with more samples have greater influence. Then the information gain of feature x_i for sample set X can be expressed as:

$$Gain(X, x_i) = Ent(X) - \sum_{v=1}^{V} w_n Ent(X^v) \quad (9)$$

If the node of feature x_i in the m tree is in the set M, then the importance of feature x_i in the m tree is

$$Gain_{mi} = \sum_{m \in M} Gain(X, x_i) \quad (10)$$

Assumed that the decision tree optimized by Lagrange multiplier method has a total of t trees, then the feature importance of feature x_i is

$$Gain_i = \sum_{m=1}^{t} Gain_{mi} \quad (11)$$

3.2 Decision tree algorithm

The characteristics of industrial Internet network attack behavior are characterized by high complexity, high dimension and data imbalance. Based on the characterization modeling of industrial control behavior in industrial situation security and the above optimization algorithm, the optimization decision tree feature processing algorithm is designed. Attackers launch attacks on factories, devices, and networks in industrial scenarios, collect and summarize attack data. The original data as input of the algorithm for training and testing, and the feature analysis results and attack classification are output. The optimization decision tree algorithm can be divided into two parts, the first part is the decision tree feature processing, the second part is the optimization loss, feature processing again. Because the optimized decision tree algorithm only needs to calculate the best value of the first order partial differential when calculating the prediction, it can obtain the global best advantage, effectively improve the calculation speed, and then improve the time ductility of the prediction. The framework of feature processing in this paper are shown in Figure 1.

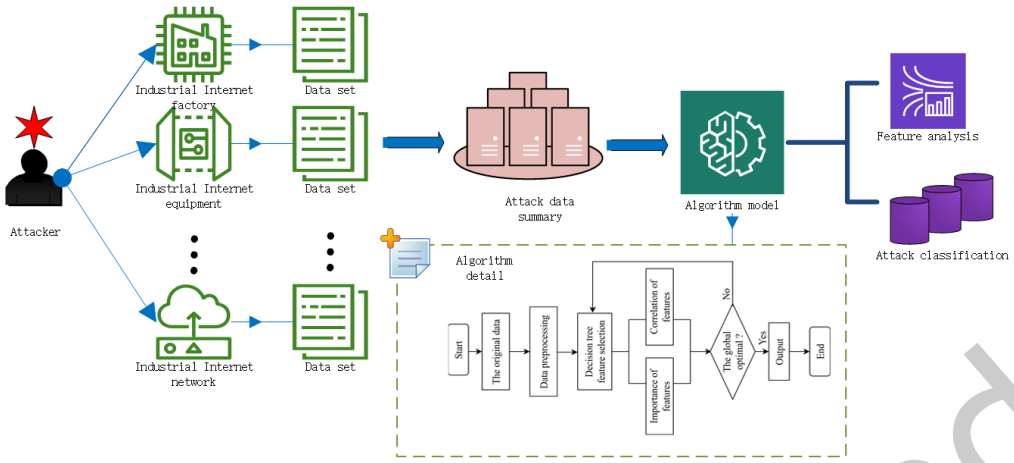


Figure 1: Feature processing framework.

ALGORITHM 1: Decision Tree Learning Algorithm Optimized by Lagrange Multiplier Method

Input : x, y;

Output: ξ

1: while *features _ list*:

2: Train DecisionTree(x, y) on *features _ list* :

3 : *RecordLoss* :

4 : Get *ranked _ feature _ importances*;

5 : del *ranked _ feature _ importances*$[-1]$;

6 : *features _ list = ranked _ feature _ importances*;

7 : end while

8 : $\xi = \arg\min_{\xi}(7)$

4 EXPERIMENTAL DATA PREPROCESSING

4.1 One-Hot Encoding

Network attack behavior in Industrial Internet contains a great number of feature categories, most of which cover industrial equipment information, network status, etc. Some features represented by equipment firmware information are expressed in the form of String, which needs to be transcoded when the algorithm is running, which will increase the complexity and cost of subsequent experiments. In this paper, one-hot encoding is used to pre-process the data of features with feature type nominal in this data set. One-hot encoding refers to the use of N-bit state registers to achieve the

encoding of N states, fully ensuring that each state can be stored in the register, no feature loss occurs, and only one bit of coding is valid.

4.2 UNSW-NB15 Data set

The Australian Cyber Security Centre (ACSC) released the University of New South Wales 2015 Network Benchmark (UNSW-NB15) data sets in 2015, which is the simulation data of the industrial intrusion detection system, and is more suitable for the Industrial Internet situational security experiment. It consists of 2,540,044 samples and this included 175341 training samples and 82332 test samples. There are 49 attack features including traffic feature, basic feature, content feature, time feature, general feature, connection feature and mark feature, and 9 different types of network attack. For the convenience of subsequent experiments, the features and attack category are respectively numbered in this paper, as shown in Table 2 and Table 3.

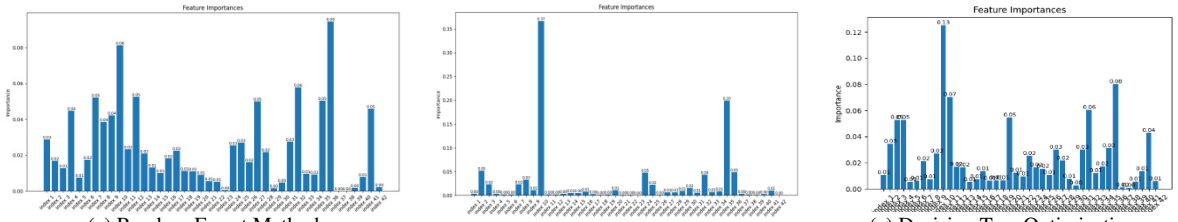
Table 2: The feature of UNSW-NB15 data set

Index	Feature	Index	Feature	Index	Feature	Index	Feature	Index	Feature	Index	Feature
1	dur	8	dbytes	15	dloss	22	dtcpb	29	trans_depth	36	ct_dst_src_ltm
2	xProt	9	rate	16	sinpkt	23	dwin	30	response_body_len	37	is_ftp_login
3	xServ	10	sttl	17	dinpkt	24	tcprtt	31	ct_srv_src	38	ct_ftp_cmd
4	xState	11	dttl	18	sjit	25	synack	32	ct_state_ttl	39	ct_flw_http_mthd
5	spkts	12	sload	19	djit	26	ackdat	33	ct_dst_ltm	40	ct_src_ltm
6	dpkts	13	dload	20	swin	27	smean	34	ct_src_dport_ltm	41	ct_srv_dst
7	sbytes	14	sloss	21	stcpb	28	dmean	35	ct_dst_sport_ltm	42	is_sm_ips_ports

Table 3: The attack type of UNSW-NB15 data set

Attack Number	Attack Type	Attack Number	Attack Type	Attack Number	Attack Type
1	Exploits	4	Reconnais	7	Worms
2	DoS	5	Shellcode	8	Backdoors
3	Fuzzers	6	Analysis	9	Generic

In this paper, the random forest method and XG-Boost method are firstly used to preliminarily screen the original features and remove irrelevant features. According to the 42 features in the UNSW-NB15 data set, the extreme points of the loss function are obtained according to the above algorithm, and the Lagrange multiplier method is used to optimize the decision tree to rank the feature importance. Before the start of all experiments, the paper normalized the experimental data to make sure the specific values of all features were limited to between [0,1]. The experiment uses the SK-learn package in Python. To ensure the credibility of the experimental results, the above two experiments have the same configuration except for the method. According to the experience of previous scholars [33], n_estimators in the random forest feature importance screening experiment were selected as 500. Feature screening results are shown in Figure 2. In order to maintain model accuracy and moderate complexity, this paper selects features with importance greater than 0.02 for subsequent model experiments.



(a) Random Forest Method (b) XG-Boost (c) Decision Tree Optimization

Figure 2: Rank of importance of feature screening.

Table 4: Feature screening of UNSW-NB15 data set

Random Forest				XG-Boost		Decision Tree Optimization
dur	rate	dinpkt	ct_srv_src	xProt	synack	xServ
xPort	sttl	tcprtt	ct_state_ttl	xServ	response_body_len	xState
xState	dttl	synack	ct_dst_sport_ltm	sbytes	ct_state_ttl	sttl
dpkts	sload	ackdat	ct_dst_src_ltm	dbytes	ct_dst_sport_ltm	dttl
sbytes	dload	smean	ct_srv_dst	sttl	ct_dst_src_ltm	swin
dbytes	sinpkt	dmean		tcprtt		

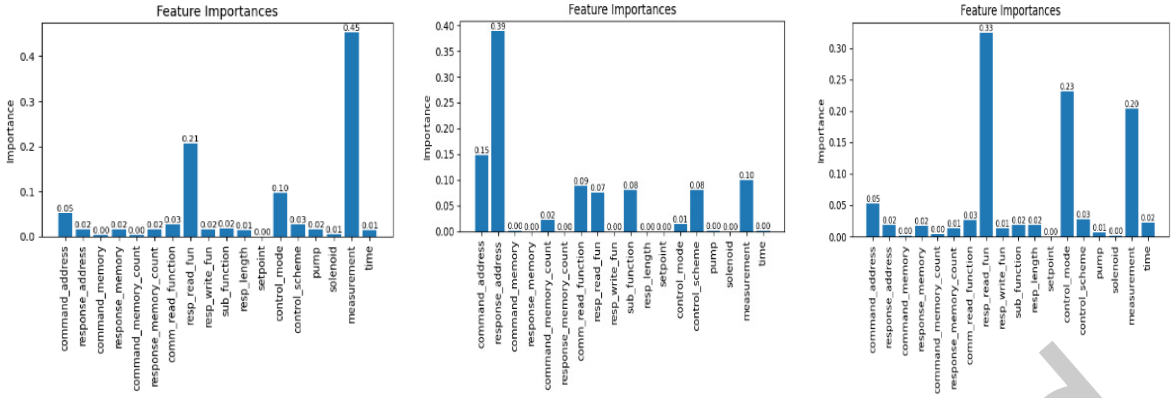
4.3 Natural Gas Pipeline Data set

In order to prove the effectiveness and portability of the proposed method, a natural gas pipeline security standard data set collected by Mississippi State University is selected for experimental verification. This data set has a total of more than 60,000 groups, including 18 features and 7 attack categories. To facilitate the demonstration of subsequent experimental results, different features and attack categories are numbered in this paper. The details are shown in Table 5.

Table 5: The type of Natural Gas Pipeline data set

index	Feature	index	Feature	Number	Attack type
1	command_address	10	sub_function	Normal	Normal
2	response_address	11	resp_length	Attack1	NMRI
3	command_memory	12	setpoint	Attack 2	CMRI
4	response_memory	13	control_mode	Attack 3	MSCI
5	command_memory_count	14	control_scheme	Attack4	MPCI
6	response_memory_count	15	pump	Attack5	MFCI
7	comm_read_function	16	solenoid	Attack6	DoS
8	resp_read_fun	17	measurement	Attack7	Recon
9	resp_write_fun	18	time		

The features of the natural gas pipeline data set were preliminarily screened, and the random forest and XG-Boost methods were also selected. In order to ensure the comparability of the results, the experiment in this part adopted the same experimental environment and parameter settings as the above experiments. In order to maintain model accuracy and moderate complexity, this paper selects features with importance greater than 0.02 for subsequent model experiments.



(a) Random Forest (b) XG-Boost (c) Decision Tree Optimization

Figure 3: Rank of importance of feature screening.

Table 6: Feature screening of Natural Gas Pipeline data set

Random Forest			XG-Boost				Decision Tree Optimization	
command_address	resp_read_fun	control_scheme	command_address	comm_read_function	sub_function	measurement	command_address	control_mode
comm_read_function	control_mode	measurement	response_address	resp_read_fun	control_scheme		resp_read_fun	measurement

5 ANALYSIS OF EXPERIMENTAL RESULTS

According to the analysis of the characteristic behavior of Industrial Internet network attacks, based on the above research, this part firstly uses K-Nearest Neighbor (KNN) algorithm to carry out the binary classification and multi-classification experiments of attack behaviors. The main feature processing method is to train and test the classification model by adding the features screened by random forest and XG-Boost one by one, and compare the performance of the model with the feature combination optimized by decision tree. At the same time, in order to verify the universality of the optimization decision tree method, different classification models are selected for performance test.

In this paper, the main indicators used to evaluate the performance of the model include the Time spent in the experiment (t) and the Accuracy of experiment with different feature detection. The value of accuracy is between [0,1].

5.1 Experiment using UNSW-NB15 data set

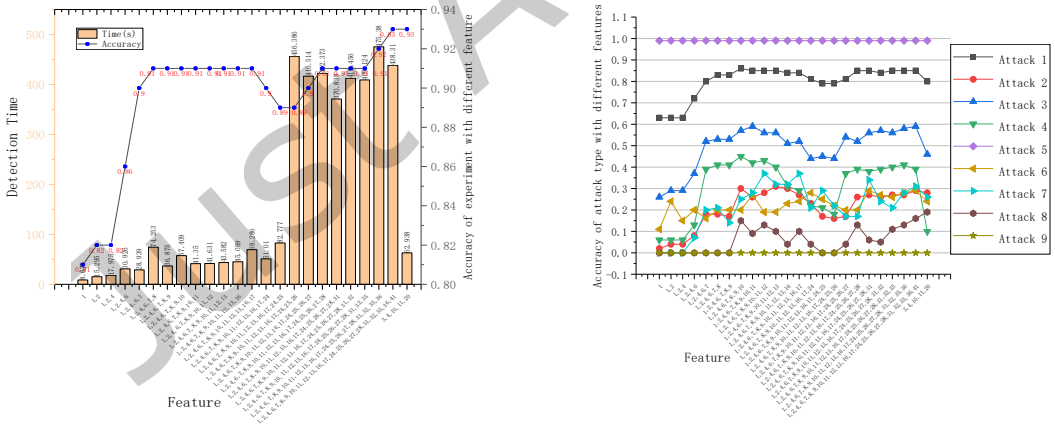
5.1.1 Experimental analysis based on random forest feature importance screening

According to the order in Table 2, different features are selected and added in turn. The experimental results are shown in Figure 4. As can be seen from Figure 4 (a), When index 26 was added to the experiment, the detection time increased significantly and remained high thereafter. However, the experimental detection accuracy improved significantly after the addition of index6, index7, index8, and then kept at a higher range. When index 24 was added, the experimental detection accuracy decreased slightly, but the detection time did not decrease accordingly. This phenomenon indicates

that index 24 is contrary to some of the aforementioned features, and the features cancel each other after addition. According to the Fig. 4, when index 36 was added to the test, the detection accuracy increased slightly and reached the maximum after index 41 was added (Accuracy was 0.93). In comparison, the experimental results of optimized feature combination reached a high level in terms of detection accuracy (consistent with the highest accuracy achieved in previous experiments, which was 0.93).

The number of optimized feature combinations is 5, and compared with the previous experiment with the same number of features, the detection time increases nearly 3 times, indicating that the optimized feature combination does not have a great advantage in terms of time consumption. However, compared with the feature combination with the same accuracy, the time consumption is reduced by about 7 times, that is, the efficiency can be significantly improved while the detection accuracy is guaranteed.

In Figure 4 (b), when the less number of features in the experiments, the Attack category of testing the overall effect is poorer, except the Attack 1 (Exploits) and Attack 5 (Shellcode), the overall low level, as the characteristic gradually after joining, Attack category test basically steady rise stage. However, after index 17(dinpkt), index 24(tcprtt), index 25(synack) and index 26(ackdat) are added, the detection accuracy drops briefly, which indicates that the above four features do not significantly help the detection accuracy of attack categories to a certain extent. In other words, these features are redundant with the aforementioned features, and the combination of them not only does not enhance the experimental effect, but also reduces the detection accuracy of attack feature categories. It can be observed that the overall experimental effect of the optimized feature combination is at a medium level, but compared with the experimental effect of adding the same number of features, it is still at a relatively good level.



(a)Accuracy and detection time of binary detection

(b)Accuracy of multi-classification detection

Figure 4: Experimental performance of different features.

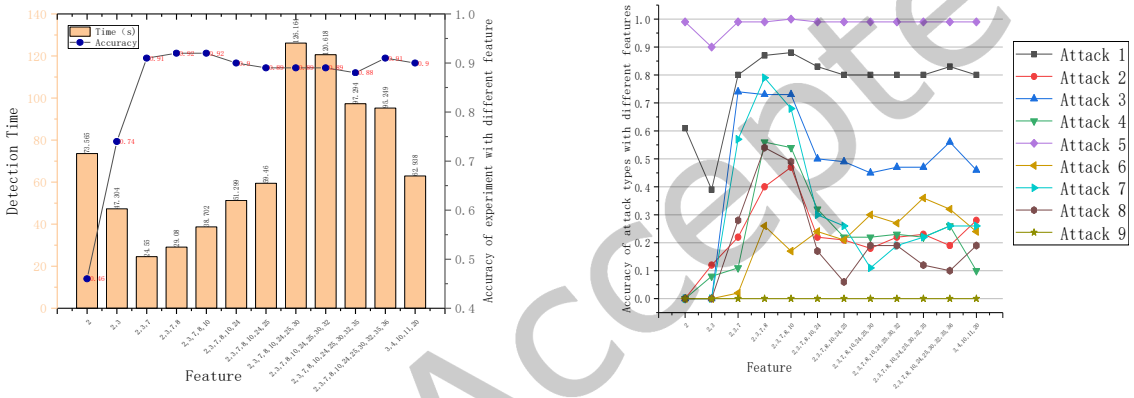
In this part of the experiment, when the number of trained features reached the maximum, the Accuracy of the experiment achieved a relatively ideal effect, which was significantly improved

compared with [9,28,31], indicating the effectiveness of preliminary screening of data features in the experiment. Moreover, the experimental accuracy, experimental time consumption and attack category detection accuracy of the optimized feature combination are far better than those of [9,28,31], which proves the superiority of the proposed method.

5.1.2 Experimental analysis based on XG-Boost feature importance screening

To maintain the validity of the experiment, we performed comparative tests again using features screened by method XG-Boost. The experimental results are as follows. The addition of index 7, index 8 and index 9 greatly improved the test accuracy of attack categories in the experiment, and basically reached the peak value at one time.

As shown in Figure 5, index 35 (ct_dst_sport_ltm) and index 36 (ct_dst_src_ltm) have a relatively good impact on the experimental results.



(a) Accuracy and detection time of binary classification detection

(b) Accuracy of multi-classification detection

Figure 5: Experimental performance of different features.

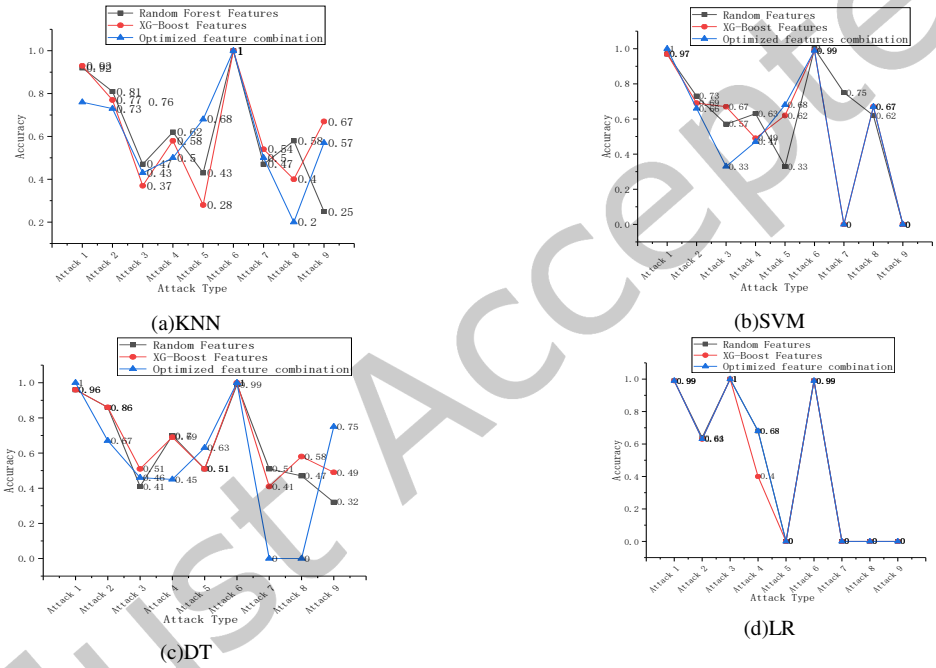
5.1.3 Performance comparison of multi-classification methods

In order to further verify the performance of the optimized decision tree algorithm, K-nearest neighbor algorithm (KNN), support vector machine (SVM), decision tree (DT) and logistic regression (LR) were selected for multi-classification detection of attack behavior, the result is shown in Figure 6.

About Attack1 (Exploits) and Attack6 (Analysis) has higher precision, and to optimize the decision tree characteristics after the combination also has better detection performance, from a certain extent, also illustrates the article the effectiveness of the proposed method.

Table 7: The performance of different model on UNSW-NB15 data set

Attack Type	Detection rate											
	KNN			SVM			DT			LR		
	RF	XG	OP	RF	XG	OP	RF	XG	OP	RF	XG	OP
Exploits	0.92	0.93	0.76	0.97	0.97	1	0.96	0.96	1	0.99	0.99	0.99
DoS	0.81	0.77	0.73	0.73	0.69	0.66	0.86	0.86	0.67	0.64	0.63	0.63
Fuzzers	0.47	0.37	0.43	0.57	0.67	0.33	0.41	0.51	0.46	1	1	1
Reconnais	0.62	0.58	0.5	0.63	0.49	0.47	0.7	0.69	0.45	0.68	0.4	0.68
Shellcode	0.43	0.28	0.68	0.33	0.62	0.68	0.51	0.51	0.63	0	0	0
Analysis	1	1	1	1	0.99	0.99	0.99	1	1	0.99	0.99	0.99
Worms	0.47	0.54	0	0.75	0	0	0.51	0.41	0	0	0	0
Backdoors	0.58	0.4	0	0.62	0.67	0	0.47	0.58	0	0	0	0
Generic	0.25	0.67	0.57	0	0	0	0.32	0.49	0.75	0	0	0



(a)KNN (b)SVM

(c)DT (d)LR

Figure 6: Effects of different methods on attack type identification.

Table 8: The performance of different schemes on UNSW-NB15 data set

Comparative literature	Exploits	Dos	Fuzzers	Reconnais	Shellcode	Analysis	Worms	Backdoors	Generic
J. Zhou [5]	76.31	35.47	63.50	92.89	91.23	70.52	82.31	72.65	84.27
J. Ren[11]	0.51	0.35	0.48	0.42	0	0.53	0.30	0	0.99
Z. Zeng [31]	0.90	0.23	0.86	0.80	1	0.98	0	0	0.98
X. Zhang [34]	0.87	0.02	0.57	0	0	0.01	0.11	0	0.84
OP	1	0.67	0.46	0.45	0.63	1	0	0	0.75

5.2 Experiment using Natural gas pipeline data set

5.2.1 Experimental analysis based on random forest feature importance screening

According to the feature screening results in Table 6, the feature combination after random forest and XG-Boost screening are compared in performance with the feature combination after optimized decision tree. Figure 7 shows the respectively based on random forest method combined with the characteristics of XG-Boost method on different kinds of attack detection performance. After optimization, the performance of the feature combination experiment is excellent, especially the f1-score is better than any of the above feature combination. As we can see, after index 2(response_address), index 8(resp_read_fun), index 10(sub_function) and index 14(control_scheme) are added, the detection accuracy is greatly improved and maintained at a high level. As we can see, after index 2(response_address), index 8(resp_read_fun), index 10(sub_function) and index 14(control_scheme) are added, the detection accuracy is greatly improved and maintained at a high level. The above four features have a strong correlation on the attack detection of the natural gas pipeline data set. Thus, the superiority of the proposed model is proved.

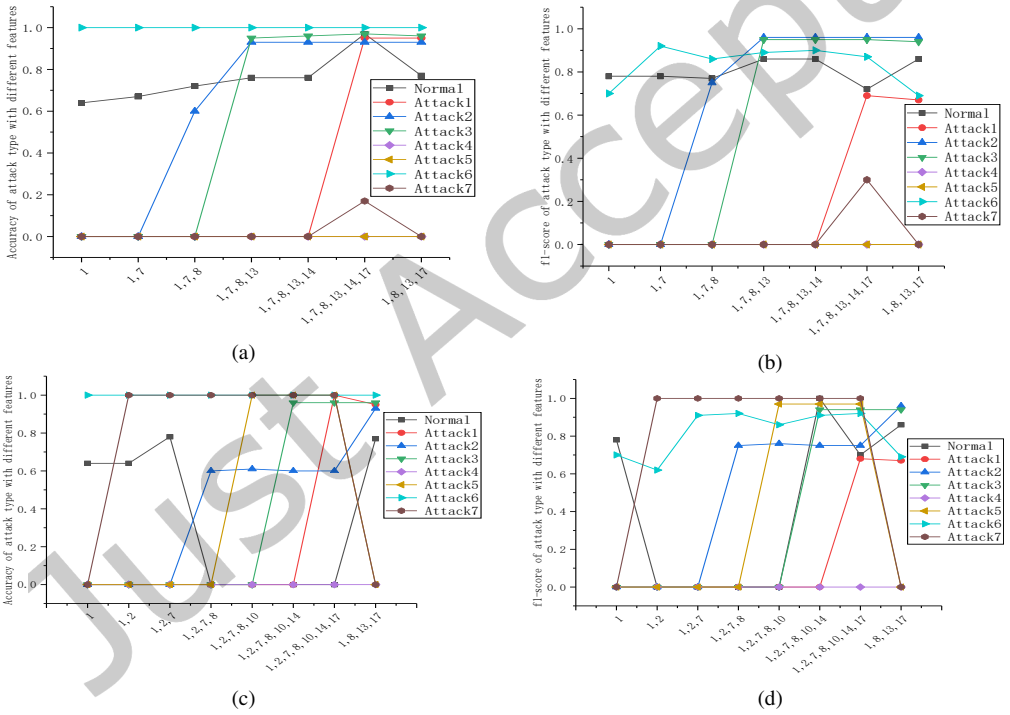


Figure 7: Effects of different methods on attack type identification.

Table 9: Accuracy of attack type detection based on random forest

Attack Features	1	1,7	1,7,8	1,7,8,13	1,7,8,13,14	1,7,8,13,14,17	1,8,13,17
Normal	0.64	0.67	0.72	0.76	0.76	0.97	0.77
NMRI	0	0	0	0	0	0.95	0.95
CMRI	0	0	0.6	0.93	0.93	0.93	0.93
MSCI	0	0	0	0.95	0.96	0.97	0.96
MPCI	0	0	0	0	0	0	0
MFCI	0	0	0	0	0	0	0
DoS	1	1	1	1	1	1	1
Recon	0	0	0	0	0	0.17	0

Table 10: Accuracy of attack type detection based on XG-Boost

Attack Features	1	1,2	1,2,7	1,2,7,8	1,2,7,8,10	1,2,7,8,10,14	1,2,7,8, 10,14,17	1,8,13, 17
Normal	0.64	0.64	0.78	0	0	0	0	0.77
NMRI	0	0	0	0	0	0	1	0.95
CMRI	0	0	0	0.6	0.61	0.6	0.6	0.93
MSCI	0	0	0	0	0	0.96	0.96	0.96
MPCI	0	0	0	0	0	0	0	0
MFCI	0	0	0	0	1	1	1	0
DoS	1	1	1	1	1	1	1	1
Recon	0	1	1	1	1	1	1	0

5.2.2Performance comparison of multiple classification methods

In this part, KNN, SVM, DT and LR are used to detect the category of attack behavior on the natural gas pipeline data set. The experimental results are shown as Figure 8. The classification effect of this data set can achieve superior performance on SVM algorithm. Among them, the detection effect of Attack 1 (NMRI) and Attack 6 (DoS) is in the leading position, and on the whole, the feature combination after XG-Boost and optimized decision tree maintains superior attack category detection accuracy.

Table 11: Attack type identification accuracy for different methods

Attack Type	Detection rate											
	KNN			SVM			DT			LR		
	RF	XG	OP	RF	XG	OP	RF	XG	OP	RF	XG	OP
Normal	0.78	0.85	0.77	0.76	0.82	0.74	0.77	0.84	0.75	0.76	0.82	0.75
NMRI	0.95	1	1	1	1	1	1	1	1	1	1	1
CMRI	0.94	0.58	0.94	0.93	0.59	0.93	0.93	0.57	0.94	0.9	0.59	0.9
MSCI	0.96	0.94	0.96	1	0.97	0	0.92	0.97	0.96	0.96	0.96	0
MPCI	0	0	0	0	0	0	0	0	0	0	0	0
MFCI	0	1	0	0	1	0	0	1	0	0	1	0
DoS	1	1	1	0.97	1	1	1	1	1	1	1	1
Recon	0	1	0.5	0	1	0	0	1	0	0	1	0

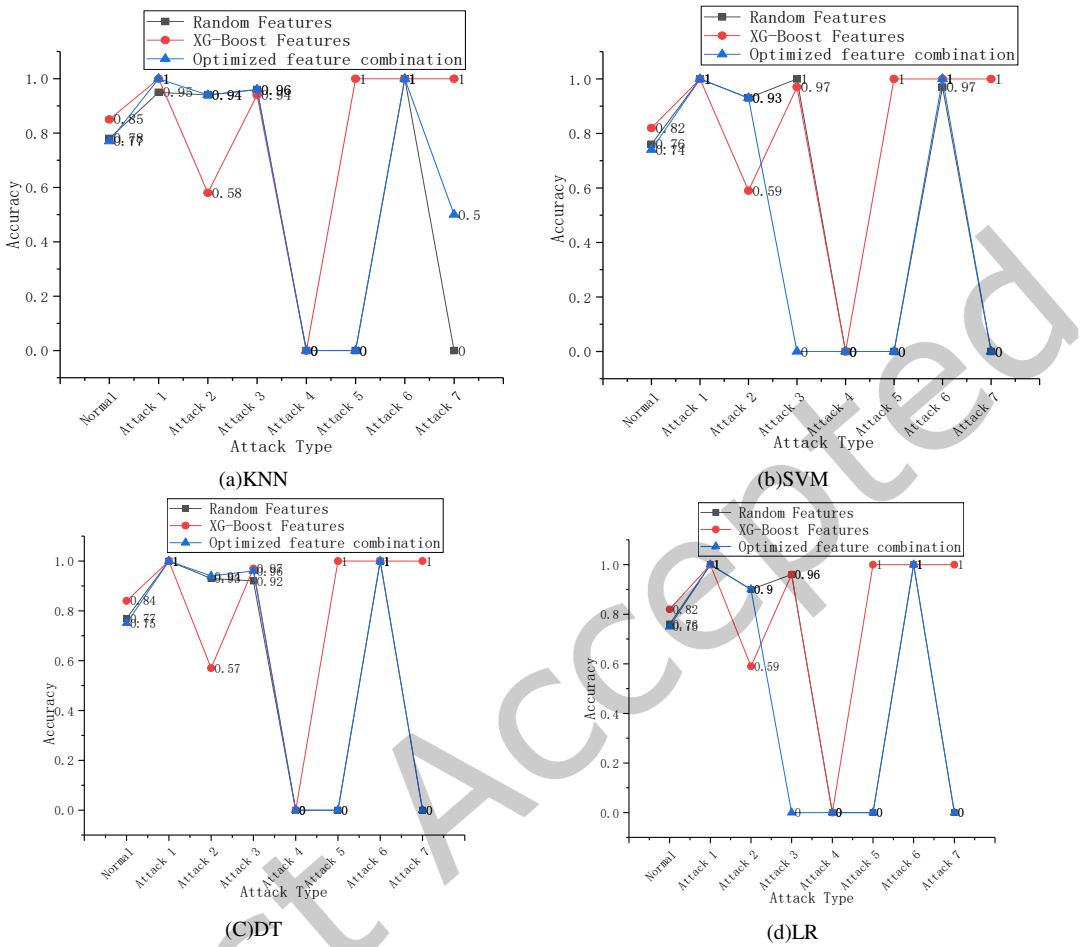
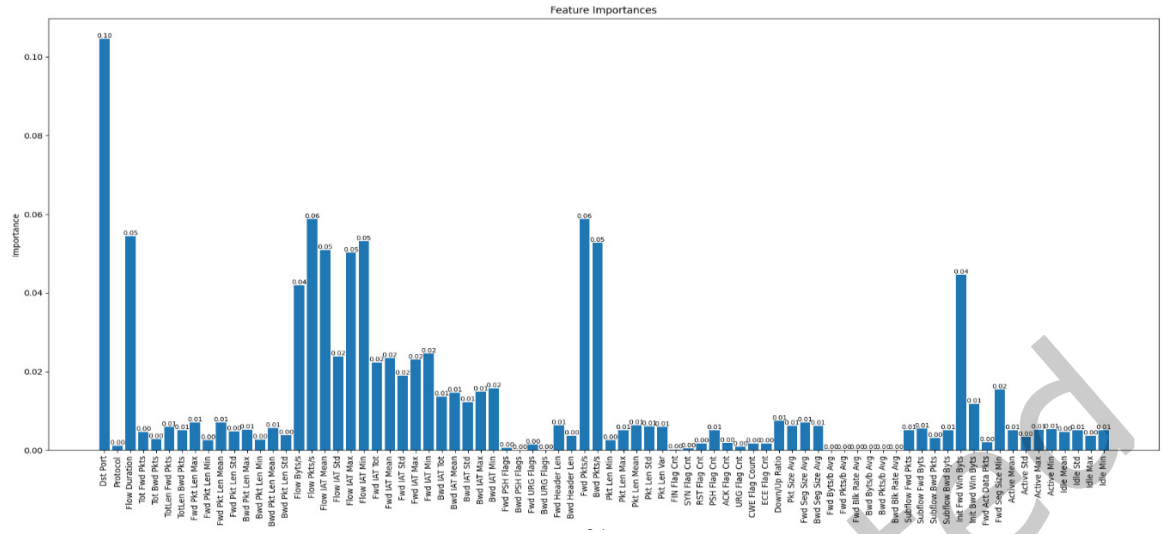


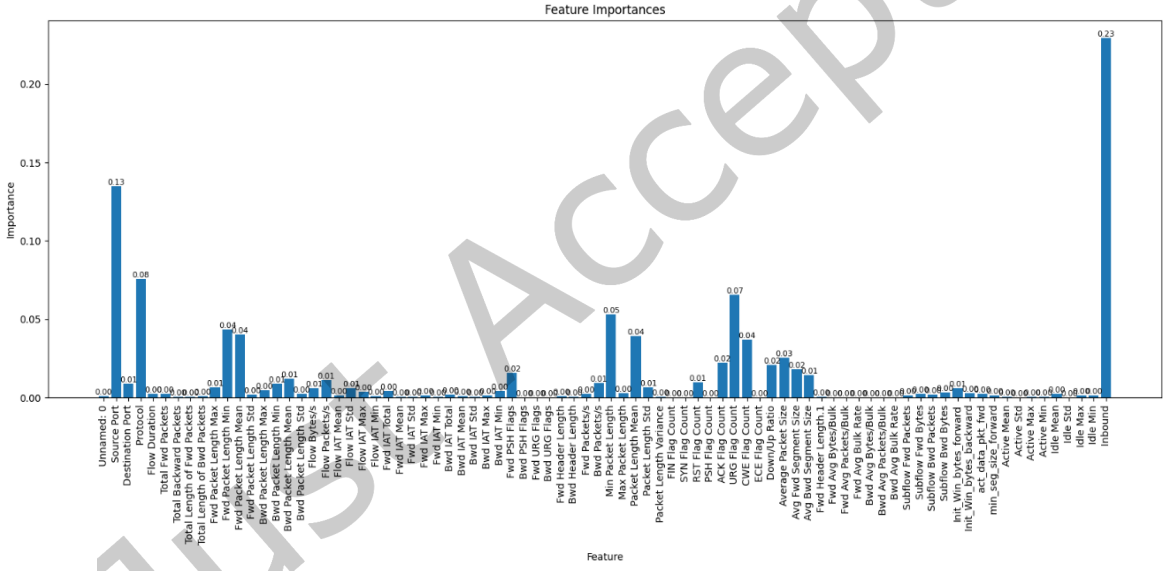
Figure 8: Effects of different methods on attack category identification.

5.3 Generalization proof

In order to prove the generalization of the proposed method, the novel CICDDoS2019 attack dataset and CICIDS2018 attack dataset in industrial scenarios are selected to prove the generalization. The CICDDoS2019 and CICIDS2018 datasets contain 82 and 78 features, respectively. The algorithm proposed in this paper is used to screen the features, and their importance ranking is shown in Figure 9.



(a)of CICIDS2018 dataset



(b)of CICDDoS2019 dataset

Figure 9: Importance ranking of feature screening.

The filtered features and all original features are used as the input of the model for attack classification detection, and the operation time of the model is recorded. The experimental results are shown in Table 12. With the same accuracy, the proposed method greatly reduces the model time, consistent with the foregoing conclusion. It shows that the proposed model has strong generalization.

Table 12: Attack type identification accuracy for different features

dataset	Features type	label	precision	recall	time-consuming
CICDDoS2019	Filtered	Attack	0.99	0.99	13.00s
		Normal	1	1	
	Full	Attack	0.99	0.99	63.46s
		Normal	0.98	0.90	
CICIDS2018	Filtered	Attack	0.8	0.92	7.05s
		Normal	0.51	0.27	
	Full	Attack	0.8	0.7	166.00s
		Normal	0.35	0.27	

5.4 Discussion

In this paper, an industrial Internet feature selection algorithm based on Lagrange multiplier method is proposed, and four different industrial security data sets are used for experimental verification. After experimental verification, it is found that the proposed method has obvious improvement in detection accuracy and detection time. The characteristics of industrial Internet network attack behavior have a very important influence on the detection accuracy of attack results. The data combination after screening by the proposed method has relatively good performance of attack category detection. Experimental results show that the proposed algorithm can reduce the time complexity while ensuring the prediction accuracy. Through the verification of four data sets, this paper shows that the algorithm has strong generalization, that is, the method can be extended to other data sets. However, it is worth noting that there are serious imbalance and functional redundancy in the sample features. For example, in the UNSW-NB15 data set, the Attack 9 (Generic) cannot be detected when the number of instances is small, so the detection accuracy of this attack category is 0. In the real network attack detection process, certain types of attack detection of low, although less classification instance does not affect the overall detection rate (for example, the paper for UNSW-NB15 data set of Attack 9 detection accuracy is 0, but does not affect the overall accuracy, other types of aggression still have higher accuracy). However, these problems can be exploited by attackers to avoid the detection system to cause damage to the network. The subsequent research will focus on the problem of uneven data samples in the detection of Industrial Internet network attack behavior.

6 CONCLUSION

The attack characteristics of industrial Internet include a great deal of industrial information including equipment information and network status, etc. This paper starts with the basic characteristics of industrial Internet network attack behavior, alleviates the serious unbalance of attack data through algorithm improvement, and analyzes the influence of different characteristics on the final attack results. In this paper, the random forest and XG-Boost methods are used to preprocess the attack behavior characteristics of the industrial security UNSW-NB15 data set and the gas pipeline security data set, the screening results are combined with the characteristics of the optimized decision tree to detect the attack behavior and attack categories of industrial Internet. Usually, the prediction accuracy and time complexity of the algorithm need to be balanced. The proposed method can effectively reduce the time complexity while taking the accuracy into account. The results show that the

optimized feature combination has higher detection accuracy and less detection time for attack results, and the precision test performance is still high after multi-classification model testing, which can prove the effectiveness and portability of the proposed method. The experimental results show that the influence degrees of the different features associated with the attack behavior can result in the binary classification attack detection increases to 0.93, and the attack detection time reduces by 6.96 times. The overall accuracy of multi-classification attack detection is also observed to improve by 0.11. In the above two data sets, nine key features of attack behavior analysis were found. The method proposed in this paper is more suitable for real-time detection of industrial Internet network attacks. The experimental results of. CICDDoS2019 dataset and CICIDS2018 dataset show that the proposed method has good generalization and can be extended to the same type of industrial anomaly data sets.

In the following research, we will focus on the imbalance of industrial Internet data and the heterogeneity of data, and further optimize the algorithm to alleviate the impact of feature imbalance on the final result. At the same time, the anomaly detection algorithm of industrial Internet is optimized to realize the synchronous improvement of feature optimization and detection performance.

ACKNOWLEDGMENTS

This work was supported by Project of Leading Talents in Science and Technology Innovation in Henan Province under Grant 204200510021, Program for Henan Province Key Science and Technology under Grant 222102210177, Grant 222102210072 and Grant 212102210383, as well as Henan Province University Key Scientific Research Project under Grant 23A520008. The work of K.-K. R. Choo was supported only by the Cloud Technology Endowed Professorship.

REFERENCES

[1] H. Shen, J. Liu, K. Chen, J. Liu and S. Moyer. 2015. SCPS: A Social-Aware Distributed Cyber-Physical Human-Centric Search Engine. *IEEE Trans. Comput*. 64, 2 (Feb. 2015), 518-532.

[2] Daniel Angermeier, Hannah Wester, Kristian Beilke, Gerhard Hansch, and Jörn Eichler. 2023. Security Risk Assessments: Modeling and Risk Level Propagation. *ACM Trans. Cyber-Phys. Syst*. 7, 1(January 2023), 25 pages. https://doi.org/10.1145/3569458.

[3] J. Li, Z. Wang, Y. Shen and L. Xie. 2022. Security Synthesis for Cyber–Physical Systems. I*EEE Trans. Syst., Man. Cybern., Syst*., early access, https://doi.org/10.1109/TSMC.2022.3189175.

[4] M. P. R. S. Kiran and P. Rajalakshmi. 2018. Performance Analysis of CSMA/CA and PCA for Time Critical Industrial IoT Applications. *IEEE Trans. Ind. Informat*. 14, 5(May 2018), 2281-2293. https://doi.org/10.1109/TII.2018.2802497.

[5] J Zhou, P He, R Qiu, G Chen, W Wu. 2021. Research on intrusion detection based on random forest and gradient boosting tree. *J. Software*. 32, 10(May 2021). 3254-3265. https://doi.org/10.13328/j.cnki.jos.006062.

[6] J. Liang, Z. Qin, S. Xiao, L. Ou and X. Lin. 2021. Efficient and Secure Decision Tree Classification for Cloud-Assisted Online Diagnosis Services. *IEEE Trans. Depend. Secure Computat*. 18, 4(July-Aug. 2021). 1632-1644. https://doi.org/10.1109/TDSC.2019.2922958.

[7] Anas Alsoliman, Giulio Rigoni, Davide Callegaro, Marco Levorato, Cristina M. Pinotti, and Mauro Conti. 2023. Intrusion Detection Framework for Invasive FPV Drones Using Video Streaming Characteristics. *ACM Trans. Cyber-Phys. Syst.* 7, 2(April 2023), 29 pages. https://doi.org/10.1145/3579999.

[8] S. Ponomarev and T. Atkison. 2016. Industrial Control System Network Intrusion Detection by Telemetry Analysis. *IEEE Trans. Dependable Secure. Comput.* 13, 2(Mar-Apr 2016). 252-260. https://doi.org/10.1109/TDSC.2015.2443793.

[9] K. Yang, Y. Shi, Z. Yu, Q. Yang, A. K. Sangaiah and H. Zeng. 2022. Stacked One-Class Broad Learning System for Intrusion Detection in Industry 4.0. *IEEE Trans. Ind. Informat.* https://doi.org/10.1109/TII.2022.3157727.

[10] H. Yao, P. Gao, P. Zhang, J. Wang, C. Jiang and L. Lu. 2019. Hybrid Intrusion Detection System for Edge-Based IIoT Relying on Machine-Learning-Aided Detection. *IEEE Netw*. 33, 5(Sept.-Oct. 2019). 75-81. https://doi.org/10.1109/MNET.001.1800479.

[11] J. Ren, Y. Zhang, B. Zhang, and S. Li. 2022. Classification Algorithm of Industrial Internet Intrusion Detection Basedon Feature processing. *J. Comp. Research. Development*. 59, 5 (Dec. 2022). 1148-1159. https://doi.org/10.7544/issn1000-1239.20211152.

[12] G. Xiong, T. S. Tamir, Z. Shen, X. Shang, H. Wu and F. -Y. Wang. 2022. A Survey on Social Manufacturing: A Paradigm Shift for Smart Prosumers. *IEEE Trans. Computat. Social. Syst*. early access. https://doi.org/10.1109/TCSS.2022.3180201.

[13] B. Wang, P. Zheng, Y. Yin, Albert Shih and L. Wang. 2022. Toward human-centric smart manufacturing: A human-cyber-physical systems (HCPS) perspective. *J. Manuf. Syst*. 63(Apr. 2022). 471-490. https://doi.org/10.1016/j.jmsy.2022.05.005.

[14] S. D. D. Anton, D. Fraunholz, D. Krohmer, D. Reti, D. Schneider and H. D. Schotten. 2021. The Global State of Security in Industrial Control Systems: An Empirical Analysis of Vulnerabilities Around the World. *IEEE Inter. Things J*. 8, 24 (Dec.2021), 17525-17540. https://doi.org/10.1109/JIOT.2021.3081741.

[15] A. Wang, W. Chang, S. Chen and A. Mohaisen. 2018. Delving Into Internet DDoS Attacks by Botnets: Characterization and Analysis. *IEEE/ACM Trans, Networking*, 26, 6 (Dec. 2018). 2843-2855. https://doi.org/10.1109/TNET.2018.2874896.

[16] D. Pliatsios, P. Sarigiannidis, T. Lagkas, and A. G. Sarigiannidis, 2020. A survey on SCADA systems: secure protocols, incidents, threats and tactics. *IEEE Commun. Surv. Tutorials,* 22(3), 1942-1976. https://doi.org/10.1109/COMST.2020.2987688.

[17] H. Lee, and A. Kobsa. 2019. Confident Privacy Decision-Making in IoT Environments. *ACM Trans. Computer-Human Interaction*. 27, 1(Dec. 2019). 1-39. https://doi.org/10.1145/3364223.

[18] Novikov, D. Nazer, Y. Roman and L. Reznik. 2008. Traffic Analysis Based Identification of Attacks. *Int. J. Comput. Sci. Appl*. 5, 2(Jan. 2008). 69-88.

[19] B. Hidayanto, R. Muhammad, R. Kusumawardani and A. Syafaat. 2017. Network Intrusion Detection Systems Analysis using Frequent Item Set Mining Algorithm FP-Max and Apriori. *Proc. Comput. Sci.* 124(Dec. 2017). 751-758. https://doi.org/10.1016/j.procs.2017.12.214.

[20] F. Guo, S. Yao, N. Zhang and Y. He, 2022, XGBoost based fake data injection attack detection method for power grid, *In Proceedings of the 2nd International Conference on Electrical Engineering and Control Science*, China, IEEE. 404-407. https://doi.org/10.1109/IC2ECS57645.2022.10087960.

[21] M. Smache, A. Olivereau, T. Franco-Rondisson and A. Tria. 2019. Autonomous Detection of Synchronization Attacks in the Industrial Internet Of Things. In *Proceedings of the 2019 IEEE 38th International Performance Computing and Communications Conference*. UK. IEEE. 1-9. https://doi.org/10.1109/IPCCC47392.2019.8958717.

[22] J. Lin and L. Liu. 2019. Research on Security Detection and Data Analysis for Industrial Internet. In *Proceedings of the 2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion.* Bulgaria. IEEE. 466-470. https://doi.org/10.1109/QRS-C.2019.00089.

[23] I. H. Sarker, M.H. Furhad. and R. Nowrozy, 2021 AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions. *SN Comput. Sci*. 2, 173. https://doi.org/10.1007/s42979-021-00557-0.

[24] A. Corallo, M. Lazoi, and M. Lezzi, 2020, Cybersecurity in the context of industry 4.0: A structured classification of critical assets and business impacts. *Comput. Ind.*, 114, 103165. https://doi.org/10.1016/j.compind.2019.103165.

[25] J. B. Awotunde, C. Chakraborty, and A. E. Adeniyi, 2021, Intrusion detection in industrial internet of things network-based on deep learning model with rule-based feature selection. *Wireless commun. mobile comput.*, 1-17. https://doi.org/10.1155/2021/7154587.

[26] D. Upadhyay, J. Manero, M. Zaman and S. Sampalli, 2021, Gradient boosting feature selection with machine learning classifiers for intrusion detection on power grids. *IEEE Trans. Netw. Serv. Manag.,* 18(1), 1104-1116, https://doi.org/10.1109/TNSM.2020.3032618.

[27] L. Zhao and X. Dong. 2018. An Industrial Internet of Things Feature processing Method Based on Potential Entropy Evaluation Criteria. *IEEE Access*. 6(Aug, 2018). 4608-4617. https://doi.org/10.1109/ACCESS.2018.2800287.

[28] S. Chakraborty, A. Onuchowska, S. Samtani, W. Jank and B. Wolfram. 2021. Machine Learning for Automated Industrial IoT Attack Detection: An Efficiency-Complexity Trade-off. *ACM Trans. Manag. Informat. Syst*. 12, 4(Oct. 2021). 1-28. https://doi.org/10.1145/3460822.

[29] J. Leevy, J. Hancock, R. Zuech and T. Khoshgoftaar. 2021. Detecting cybersecurity attacks across different network features and learners. *J. Big. Data.* 8, 38(Feb. 2021). 1-29. https://doi.org/10.1186/s40537-021-00426-w.

[30] H. Binyamini, R. Bitton, M. Inokuchi, T. Yagyu, Y. Elovici and A. Shabtai. 2021. A Framework for Modeling Cyber Attack Techniques from Security Vulnerability Descriptions. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. USA. ACM. 2574–2583. https://doi.org/10.1145/3447548.3467159.

[31] Z. Zeng, B. Zhao, W. Meng and H. Chao. 2022.Towards Intelligent Attack Detection Using DNA Computing. *ACM Trans. Mult. Comput*. 19, 3s(Sep. 2022). 1-27. https://doi.org/10.1145/3561057

[32] S. Braun, S. Albrecht and S. Lucia. 2022. Attack Identification for Nonlinear Systems Based on Sparse Optimization. I*EEE Trans. Autom. Control.* 67, 12(Dec. 2022). 6397-6412. https://doi.org/10.1109/TAC.2021.3131433.

[33] Breiman. 2001. Random forests. *Machine Learning*. 45, 1(Oct. 2001). 5-32. https://doi.org/10.1023/A:1010933404324.

[34] X. Zhang, J. Li and D. Zhang. Research on feature processing for cyber attack detection in Industrial Internet of Things. In *Proceedings of the 2020 International Conference on Cyberspace Innovation of Advanced Technologies*. China. ACM. 256–262. https://doi.org/10.1145/3444370.3444581.

[35] D. Pliatsios, P. Sarigiannidis, T. Lagkas, and A. G. Sarigiannidis, 2020. A survey on SCADA systems: secure protocols, incidents, threats and tactics. *IEEE Commun. Surveys & Tutorials*, 22(3), 1942-1976. https://doi.org/ 10.1109/COMST.2020.2987688.