Hindawi
Scientific Programming
Volume 2022, Article ID 1357182, 13 pages
https://doi.org/10.1155/2022/1357182

Research Article

A Novel Method for Network Intrusion Detection

Hongmin Wang, Qiang Wei, and Yaobin Xie

State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, China

Correspondence should be addressed to Yaobin Xie; xybsoft@meac-skl.cn

Received 11 October 2022; Accepted 8 November 2022; Published 21 November 2022

Academic Editor: Cristian Mateos

Intrusion detection is one of the key research directions of network information security. In order to make up for the deficiencies of traditional security technologies such as firewall, encryption, and authentication, by analyzing the characteristics of network attacks and existing intrusion detection models, the advantages of triadic concept analysis and the application of fuzzy set theory in network intrusion detection are analyzed. The intrusion detection model FCTA based on triadic concept analysis is proposed, which promotes the further development of network intrusion detection. First, we analyze the characteristics of the data and use TF-IDF and Z-Score to normalize and standardize the data to construct a fuzzy triadic background containing quadratic characteristics. It is used to describe the triadic relationship between network connections, network connection characteristics, and intrusion types of network packets. Then, the (i)-induced operator is used to construct the fuzzy triadic concept set based on the fuzzy triadic background and transformed into a fuzzy attribute triadic concept set. Then, the new samples are classified by calculating the similarity between the new samples and the elements in the fuzzy attribute triadic concept set by using the Euclidean distance formula. In order to reduce the model space complexity, compression storage technology is adopted in the model building process.. Finally, by using the IDS-2018 dataset, the rationality and effectiveness of the FCTA model are demonstrated. The average accuracy and average intrusion detection rate of FCTA classification are much higher than BP neural network, SVM algorithm, and KNN algorithm, and the FCTA misjudgment rate is much lower than the BP neural network algorithm, the KNN algorithm, and the SVM algorithm; with the increase of data volume, the accuracy rate and intrusion detection rate increase significantly.

1. Introduction

In recent years, with the rapid development of technologies such as big data, artificial intelligence, 5G, and blockchain, physical systems and information systems have gradually achieved a high degree of integration [1], and information processes and physical processes exchange information in real time [2]. The complex integration of decision-making units and physical devices in cyberspace improves system performance while also bringing potential safety hazards. Network intrusion detection is the core of network security defense-in-depth system. Researchers at home and abroad have paid more and more attention to network intrusion detection and carried out extensive and in-depth research. Intrusion detection is to collect and analyze the network behavior of the system and use the comparison with known intrusion behavior models or the judgment analysis of unknown intrusion behavior to detect, whether there are suspicious intrusions and attacks against the system [3]. Aiming at the classification of intrusion detection, [4] proposed nonsymmetric deep autoencoder (NDAE) for unsupervised feature learning. Furthermore, authors also propose the novel deep learning classification model constructed using stacked NDAEs. Gao et al. [5] propose a support vector machine intrusion detection model (AN-SVM) based on self-encoding networks. The network connection characteristics of network connection data are extracted through data mining technology, the communication network connection data are processed into training datasets, and machine learning algorithms are used to construct the intrusion detection model [6]. At present, researchers have proposed a variety of intrusion detection methods and models based on neural networks, SVM (support vector machines), and pattern matching in data mining and machine learning algorithms. [7] made the latest investigation and research on ICS security and discussed the applicability of machine learning technology in ICS network security. [8] developed a network intrusion detection system using an unsupervised learning algorithm autoencoder and verified its performance. In [9], the CNN network is used for network intrusion detection, which achieves the purpose of detection by enhancing the feature learning ability, mainly by adding convolution kernels. In [10], an intrusion detection method based on AE-1SVM was proposed by using unsupervised learning technology to solve the problems that LightGBM cannot effectively detect unknown attacks, lack of label data in the real environment, and the high cost of manual labeling. [11] focused on network intrusion detection using convolutional neural networks (CNNs) based on LeNet-5 to classify the network threats. Wu and Guo [12] considered the existence of spatial and temporal features in the network traffic data and propose a hierarchical CNN + RNN neural network, LuNet. [13] propose a new intelligent agent-based mobile ad hoc network intrusion detection model that combines attribute selection, outlier detection, and enhanced multiclass SVM classification methods. This system detects anomalies with a low false alarm rate and a high-detection rate. . et al.[14] propose a new intrusion detection system that optimizes the number of features by developing a new feature selection algorithm based on intelligent conditional random fields (CRF), and the major advantages of this proposed system are reduction in detection time, increase in classification accuracy, and reduction in false alarm rates. In [15], a survey on intelligent techniques for feature selection and classification for intrusion detection in networks based on intelligent software agents, neural networks, genetic algorithms, neurogenetic algorithms, fuzzy techniques, rough sets, and particle swarm intelligence has been proposed. [16] propose a new feature selection algorithm called the conditional random field and linear correlation coefficient-based feature selection algorithm to select the most contributed features and classify them using the existing convolutional neural network.

Formal concept analysis [17], as an effective method for mining data associations, has been widely used in the fields of information retrieval, knowledge discovery, association analysis, recommendation systems, and software engineering [18]. In 1995, Lehmann and Wille were inspired by the philosophy of pragmatism to extend the analysis of formal concepts to the context of triadic and proposed the analysis of triadic concepts [17, 19]. In their study, Lehmann and Wille proposed the basic concepts of three element background, three element concept, and three element lattice for the first time. As an extension of formal concept analysis, triadic concept analysis is a relatively new and important research branch in the field of artificial intelligence, involving machine learning, data mining, and information retrieval [20]. [19] studied the application of triadic concept analysis in text classification. [21] studied the construction algorithm of concept triadic lattice and its application in folksonomy classification. Triadic concept analysis has a wide range of application prospects. In today's vast amount of data, it is bound to provide us with an efficient and very practical application method [20].

Triadic concept analysis can describe network connection data in a formal way and can mine the relationship between features and intrusion types, which has a good effect on classification problems. Furthermore, industrial control system intrusion detection is a classification problem, and through experiments, it is also verified that the application of ternary concept analysis in industrial control system intrusion detection has a good classification effect. Based on the above discussion, this study proposes a network intrusion detection model based on the advantages of ternary concept analysis. First, the data are processed as a fuzzy triadic background with quaternary characteristics, and a fuzzy ternary concept set containing the triadic relationship among network connection, network connection feature, and intrusion type is constructed by (i)-induction operator. Then, in order to improve the accuracy of network connection classification, fuzzy triadic concepts are transformed into fuzzy attribute concepts. Then, in order to classify the new sample, the Euclidean distance formula is used to calculate the similarity between the new sample and the element in the fuzzy attribute triadic concept set. The greater the similarity is, the more similar the new sample is to the element. Finally, the efficiency and accuracy of the model are proven by experiments.

2. Related Work

The related works are explained in the following sections.

2.1. Basic Theory of Triadic Concept Analysis. Triadic concept analysis is a high extension of formal concept analysis. It can accurately mine the effective information contained in complex cascading data, describe the mapping relationship between objects and features, use the numerical logic and arithmetic operations of operators to construct triadic concepts, and apply theoretical topology in the fields of recommendation and classification. In order to integrate the theory, the basic concepts of triadic concept analysis are given below.

Definition 1. (triadic background [17]). the triadic background is defined as a quadratic (G, M, B, and Y), where G is the set of objects, M is the set of attributes, B is the set of conditions, and Y is the triadic relationship between G, M, and B; it is $Y \in G \times M \times B$. $(g, m, b) \in Y$ indicates that the object g has attribute m under condition b.

Definition 2. (triadic concept [17]). Triadic background $K = (K_1, K_2, K_3, Y)$, for triples (A_1, A_2, A_3), $A_i \subseteq K_i$, $i = 1, 2, 3$. If $A_i = (A_j \times A_k')^{(i)}$, where $\{i, j, K\} = \{1, 2, 3\}$ and $j < k$, then (A_1, A_2, A_3) is called the triadic background K triadic concept. Among them, A_1 is called extension, A_2 is called intension, and A_3 is called mode.

Definition 3. ((*i*)-induced operator [17]). Triadic background $K=(K_1, K_2, K_3, Y)$ satisfies $j<k$ and $X\subseteq K_i$ and $Z\subseteq K_j \times K_k$, $\{i, j, k\} = \{1,2,3\}$, then (*i*)-induced operator is defined as

$$X \mapsto X^{(i)}: = \left\{\left(a_j, a_k\right) \in K_j \times K_k \middle| \left(a_i, a_j, a_k\right) \in Y, \forall a_i \in X\right\},$$
$$Z \mapsto Z^{(i)}: = \left\{a_i \in K_i \mid \left(a_i, a_j, a_k\right) \in Y, \forall \left(a_j, a_k\right) \in Z\right\}. \quad (1)$$

Definition 4. (preorder and equivalence relationship [22]). Triadic background $K=(K_1, K_2, K_3, Y)$, S is the set of all triadic concepts in the triadic background, for any $(A_1, A_2, A_3)\subseteq S$ and $(B_1, B_2, B_3)\subseteq S$. We define the preorder relationship $\lesssim i$ and the equivalent relationship $\sim i$ as

$$\left(\mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3\right) \lesssim \mathrm{i}\left(\mathrm{B}_1, \mathrm{B}_2, \mathrm{B}_3\right) \Longleftrightarrow \mathrm{A}_\mathrm{i} \subseteq \mathrm{B}_\mathrm{i},$$
$$(\mathrm{A1}, \mathrm{A2}, \mathrm{A3}) \sim \mathrm{i}(\mathrm{B1}, \mathrm{B2}, \mathrm{B3}) \Longleftrightarrow \mathrm{A}_\mathrm{i} = \mathrm{B}_\mathrm{i}, \mathrm{i} = \{1, 2, 3\}. \quad (2)$$

2.2. Basic Theory of Network Intrusion Detection. Network intrusion or attack is a multistage and step-by-step process [23, 24]. In order to find out whether there is malicious attack behavior in network access, the intrusion detection system analyzes the relevant data by monitoring the traffic in the network and the log information of the computer and finally achieves the identification of the internal personnel and the computer [25, 26]. The intrusion detection system identifies outside infiltrators who use, misuse, and abuse computer systems without authorization [27–29]. The network intrusion detection deployment is shown in Figure 1.

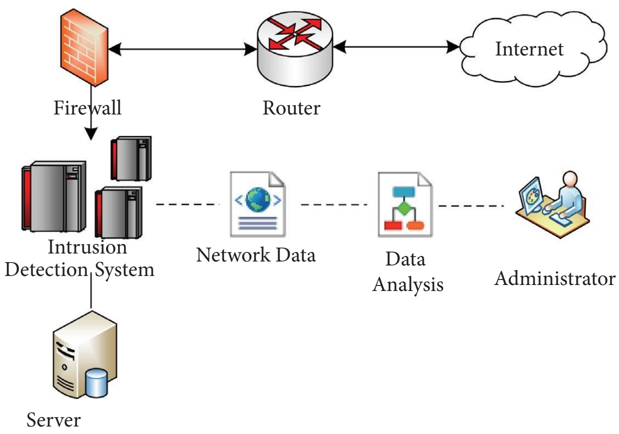

Figure 1: Network intrusion detection deployment.

The intrusion detection framework defined by DARPA [30] is shown in Figure 2. In Figure 2, the event analyzer compares the feature values extracted by the event generator from the data source with the data in the event database to determine that the data source contains abnormal data.

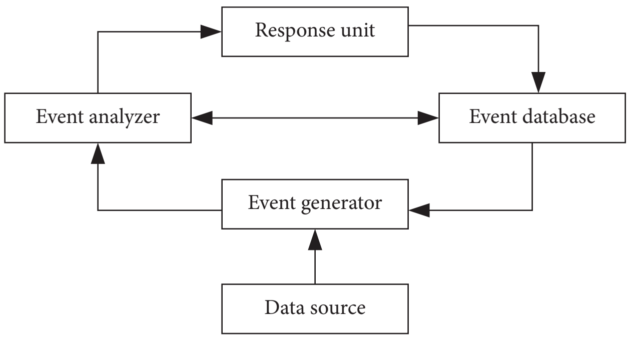

Figure 2: Intrusion detection framework.

Due to the real-time requirements of network and limited equipment resources, the existing intrusion detection for the network still has shortcomings, such as low detection efficiency and inability to effectively identify unknown attacks. Therefore, this study designs a reasonable intrusion detection based on triadic concept analysis. Methods and frameworks are proposed to improve the intrusion detection rate of the network, high rate of packet miss, false negatives, and false positives. Moreover, due to the multisource, complexity, and high-dimensional characteristics of network data, it is necessary to design reasonable data preprocessing steps. In the data preprocessing stage, this study designs data standardization formulas for different types of original data to reduce data complexity and removes invalid attributes to reduce data dimension.

3. Network Intrusion Detection Model Based on Triadic Concept Analysis

Through the previous analysis, in view of the unique advantages of ternary concept analysis, this study constructs a network intrusion detection model, and the construction process of the model is shown in Figure 3.

The detection model FCTA analyzes the characteristics of the data and uses TF-IDF and Z-Score to normalize and standardize the data and to construct a fuzzy triadic background containing quadratic characteristics. It is used to describe the triadic relationship between network connections, network connection characteristics, and intrusion types of network packets. Then, the (*i*)-induced operator is used to construct the fuzzy triadic concept set based on the fuzzy triadic background and transformed it into a fuzzy attribute triadic concept set. Then, the similarity between the new sample and the elements in the fuzzy attribute ternary concept set is calculated by Euclidean distance, determining whether the new sample is normal access data or network attack data, and if it is network attack data, we determine the attack type through the intrusion detection model.

Next, each step of the model is described in detail.

3.1. Data Preprocessing. The original data consist of records; each record contains a variety of features, including discrete and continuous. The data structure is complex, and the numerical data have different value ranges and lack normativeness. Therefore, it is necessary to normalize and standardize the original data into a unified style, which is convenient for model processing and improves efficiency. In the process of processing raw data, it represents the original

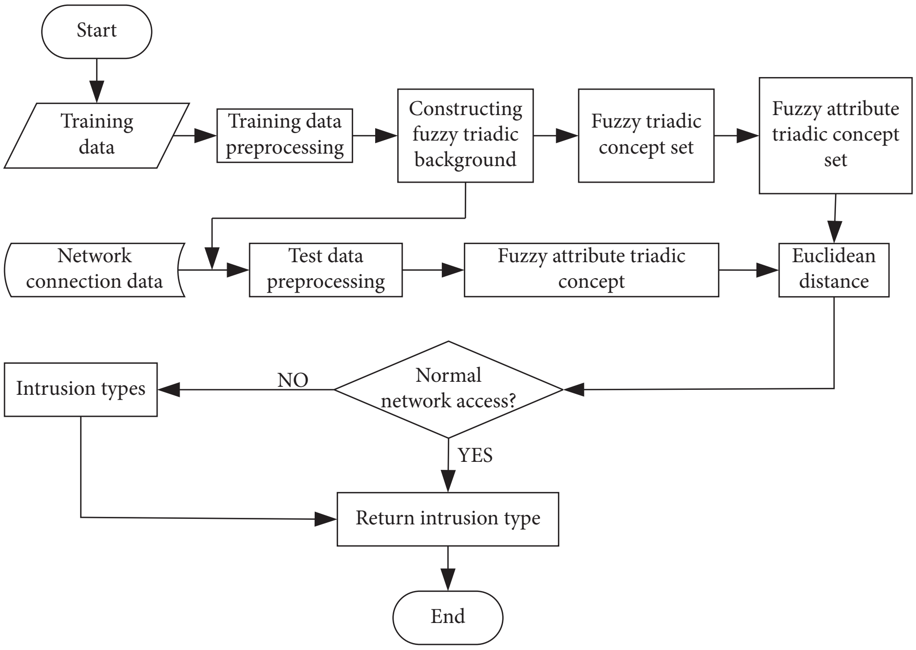
Start
Training data
Training data preprocessing
Constructing fuzzy triadic background
Fuzzy triadic concept set
Fuzzy attribute triadic concept set
Network connection data
Test data preprocessing
Fuzzy attribute triadic concept
Euclidean distance
Intrusion types
NO
Normal network access?
YES
Return intrusion type
End

FIGURE 3: Attribute class triadic concept vector intrusion detection model diagram.

data as a string type. Based on the principle of calculating feature words in the field of natural language processing, TF-IDF [31] (term frequency-inverse document frequency and word frequency-inverse document frequency) is adopted. The improved formula of algorithm normalizes the original data to a value between 0 and 1. The calculation formula is as follows:

$$w_i = \frac{f_i \times log_2 \left(n/n_i + 1 \right)}{\sqrt{\sum_j^t f_j^2 \left(log_2 \left(n/n_i + 1 \right) \right)^2}}, \tag{3}$$

where w_i represents the processed value of the ith connection feature, f_i is the frequency of the connection feature in the record, n is the amount of data in the original data, n_i is the amount of data containing the connection feature, and t is the number of connection features recorded.

Then, we use Z-Score to standardize the data so that each network connection feature value in the feature column obeys the standard normal distribution; the calculation formula is as follows:

$$x_i^* = \frac{x_i - \overline{x}}{\sigma}, \tag{4}$$

where x_i^* represents the weight of the ith network connection feature for the processed value, x_i is the eigenvalue of the ith network connection, $\overline{x}$ is the average value of the ith network connection, and σ is the standard deviation of all data in the column where the ith network connection feature is located.

3.2. Constructing Fuzzy Triadic Background and Fuzzy Triadic Concept.

In the network intrusion detection model FCTA constructed in this study, the fuzzy triple background is a quadruple (O, A, C, and R). In the quadruple, O is the network connection dataset, A is the network connection characteristic set, C is types of network intrusions set, and R is the fuzzy triadic relationship set between O, A, and C. Each fuzzy triadic relationship has a weight value of the network connection characteristic with a value range of [0, 1]. The weight value of the network connection feature can not only characterize the affiliation relationship between the attribute and the object but also represent the importance of the attribute in the object under the specific attack type. Table 1 shows the correspondence between the network connection data and the objects, attributes, and conditions in the fuzzy triadic background.

The core of the formal concept is constructed by analogy to the formal background. The fuzzy ternary concept is obtained by calculating the fuzzy ternary background through the $i - (i, j, A_k)$ induction operator [32, 33]. Under fuzzy triadic background (K_1, K_2, K_3, Y), the constructed fuzzy triadic concept is defined as a triplet (A_1, A_2, A_3); among them, LK_1, $LK_{2,}$ and LK_3 are all fuzzy sets on K_1, $K_{2,}$ and $K_{3,}$ respectively. And, for any $A_1 \in LK_1$, $A_2 \in LK_2$, and $A_3 \in LK_{3,}$ there are $\mathrm{A_i = A_j^{(i,j,A_k)}}$, $\mathrm{A_j = A_k^{(j,k,A_i)}}$, and $\mathrm{A_k = A_i^{(k,i,A_j)}}$, $\{\mathrm{i, j, k}\} = \{1, 2, 3\}$ [32]. Then, in the fuzzy triadic concept of network connection data construction, the object is a subset of the fuzzy set of network connection, the attribute is the subset of the fuzzy set of network connection characteristics, and the condition is the subset of the fuzzy set of intrusion types.

Unlike the triadic concept analysis multicondition itself, network intrusion detection is a single-classification problem, a network connection is classified as a specific intrusion type. Therefore, the condition in the constructed fuzzy triadic concept is a specific type of intrusion. For any network connection, under different intrusion types, the weights of network connection features are different. The

Table 1: Correspondence between network connection dataset and objects, attributes, and conditions in the fuzzy triadic background.

Internet connection	Blurred triadic background
Network connection packet	Objects (O)
Network connection characteristics	Attributes (A)
Types of network intrusions (Benign, Brute Force, Botnet, DDoS, DoS, infiltration, SQL injection, and Web Attacks)	Conditions (C)

network connection feature is the property of network connection, which is the concrete representation of network connection. The actual research of network intrusion detection is to determine the relationship of attack types according to the characteristics, that is, under the conditions of given network connection characteristics, to determine which type of attack the network connection belongs to. Therefore, in order to better classify network connections, it is necessary to operate in a common feature space and transform it into a fuzzy attribute ternary concept considered from the attribute point of view.

3.3. Constructing Fuzzy Attribute Triadic Concept. For any fuzzy ternary concept (O, A, and C) constructed by the network connection data, we traverse any feature A_i in the network connection feature set A, and A_i has a weight value W_i for each network connection O_i in the network connection set to O. We calculate the average weight value of A_i based on the value in the fuzzy triadic background, which indicates the degree of membership or importance of A_i. The membership degree of all network connection features to this fuzzy triadic concept is expressed as $(\overline{w_1}, \overline{w_2}, \ldots, \overline{w_\mathrm{n}})$. Therefore, transforming (A, O, and C) into a fuzzy attribute ternary concept is expressed as $\mathrm{Con} = (\{\mathrm{T}_1(\overline{w_1}), \mathrm{T}_2(\overline{w_2}), \ldots, \mathrm{T}_\mathrm{n}(\overline{w_\mathrm{n}})\}, \{O\_(1), O\_(2), \ldots, O\_(n)\}, \{C\_(k)\})$, where $\{\mathrm{T}_1(\overline{w_1}), \mathrm{T}_2(\overline{w_2}), \ldots, \mathrm{T}_\mathrm{n}(\overline{w_\mathrm{n}})\}$ is the set of network connection features with weights,$\{O_1, O_2, \ldots, O_n\}$ is the set of network connections, and C_k is a definite intrusion type. The average weight value of each network connection feature forms a vector $(\overline{w_1}, \overline{w_2}, \ldots, \overline{w_\mathrm{n}})$ and intrusion type C_k contains one or more such vectors. Therefore, the network connection can be expressed as a fuzzy attributes triadic concept vector of $V = (\overline{w_1}, \overline{w_2}, \ldots, \overline{w_\mathrm{n}})$. Then, the vector is unitized according to formula (3) to construct a fuzzy attribute triadic concept vector model. Among them, the vector unitization calculation formula is shown in formula (3):

$$V_{ij} = \frac{w_j}{\sqrt{\sum_1^n (w_k)^2}}, \tag{5}$$

where V_{ij} is the unitized value of the jth attribute vector of object i, w_j is the absolute value of the weight value of the attribute j in the fuzzy attribute ternary concept vector, w_k is the weight value of the attribute k of the object, the value of k is from 1 to n, and n is the number of attributes.

3.4. Classify Network Connections Based on the Fuzzy Attribute Triadic Concept Vector Model. New network connections are classified using the constructed fuzzy attribute triadic concept vector model. The new network connection is classified by using the training attribute class triadic concept vector model. The network connection to be classified is transformed into attribute class triadic concept vector, and the network connection is classified by comparing the similarity between vectors. The greater the similarity between vectors is, the closer the network connection is to the intrusion type, so as to determine the intrusion type of new network connection. Since the vector in the model and the vector transformed by the network connection to be classified are all unit vectors, therefore, the similarity calculation formula of two vectors is defined based on the Euclidean distance formula as

$$\mathrm{Sim}(\mathbf{C}, \mathbf{V}) = 2 - \sqrt{\sum_{i=1}^{n} (CW_i - VW_i)^2}, \tag{6}$$

where Sim(C, V) represents the calculation result, CW_i is the value of the ith attribute of the fuzzy attribute ternary concept vector constructed by the network connection to be classified, and VW_i is any one of the fuzzy attribute ternary concept vector models. where n is the number of network connection features,and i is the value of the ith attribute of the vector.

3.5. The Fuzzy Attribute Triadic Concept Vector Model Pseudocode Description. The fuzzy attribute triadic concept vector model (FCTA) mainly includes the following three steps, and the detailed process is described by Algorithms 1–3 respectively.

In step 1, *cf*Map is a dictionary class, which is designed to store the correspondence between classification and its conversion into a numerical value. The attack type is key, and the corresponding number of attack type is value. When reading the original data and converting it into a string array, the result is converted into a numeric string. Steps 2~7 are to read the original file data and converted into a two-dimensional string array. Steps 8~13 are calculate the TF-IDF value of the connection feature by formula (1) of the molecular calculation formula. We calculate the value of the molecule in the formula and then accumulate the square of each value to prepare for the data unit processing. Step 14 normalizes and standardizes the above results and calculates the ratio of TF-IDF value and sum of squares as the final result of unitization.

Step 1: we set an empty set TConSe to store the generated fuzzy triadic concepts. Steps 2 to 9 sequentially construct the

```
Input: raw data file
Output: fuzzy triadic background (FTBac)
(1)  $cf$Map ⟵ relationship between intrusion type and number
(2)  for each $i \in \{\text{raw data}\}$ do
(3)    for each $j \in \{\mathbf{T}\}$ do
(4)      strDS ⟵ Record$[i][j]$
(5)    end for
(6)    strDS ⟵ Record$[i][m]$ converted by $cf$Map
(7)  end for
(8)  for $i = 1$ to $n$ do
(9)    $cv$ ⟵ strDS$[i][m]$
(10)   for $j = 1$ to $m - 1$ do
(11)     FTBac$[i][cv][j]$ ⟵ Calculate TF – IDF value
(12)   end for
(13) end for
(14) FTBac ⟵ standardize and normalize FTBac
```

Algorithm 1: Data preprocessing algorithm.

```
Input: fuzzy triadic background (FTBac)
Output: fuzzy attribute triadic concept vector (FCTA)
(1)  Define and initialize TConSe ⟵ $\varnothing$
(2)  for each $j \in \{\mathbf{C}\}$ do
(3)    binDS ⟵ relationship between object and attribute
(4)    for $i = 1$ to $k$ do
(5)      triCon ⟵ $\left\{j, \left\{w_i | w_{ij} > 0\right\}, i\right\}$
(6)      TConSe ⟵ {triCon, triCon∩TConSe}
(7)    end for
(8)    TConSe ⟵ {TConSe|TConSe $\neq \varnothing$}
(9)  end for
(10) for $i = 1$ to TConSe.size() do
(11)   for $j = 1$ to $m$ do
(12)   $\overline{w_{ij}} \longleftarrow Avg(w_{ij})$
(13)   end for
(14) end for
(15) for $i = 1$ to TConSe.size() do
(16)   FCTA ⟵ unitize TConSe
(17) end for
(18) return FCTA
```

Algorithm 2: Algorithm of constructing the fuzzy attribute triadic concept vector model.

fuzzy triadic concept under each condition of the fuzzy triadic background. The fuzzy triadic background data are stored as a 3-dimensional array and each condition is a two-dimensional data matrix composed of objects and attributes. First, we convert the two-dimensional data matrix under the ith condition into an array list object and then loop through each row array in the array list object to construct a fuzzy ternary concept triCon. If TConSe does not contain triCon, we add triCon to TConSe. Anyway, the triCon and other fuzzy triadic concepts in TConSe are operated to generate a new fuzzy ternary concept and stored in TConSe. Steps 10 to 14 are to transform the fuzzy triadic concept set into a fuzzy attribute triadic concept vector set and take the average weight value of all extensions of attributes in the fuzzy triadic concept as the weight value of this attribute of the fuzzy triadic concept. Steps 15~17 unitize each attribute in the fuzzy attribute triadic concept vector. Step 18 returns the fuzzy attribute triadic concept vector model.

Based on the fuzzy attribute triadic concept vector model, the pseudocode of the algorithm for classifying network connections by formula (4) is described as follows.

Step 1 relies on formulas (1) and (2) to preprocess the network connection I data to obtain a fuzzy triadic background. Steps 2 to 3 are to construct a triadic concept for the network connection, convert it into a triadic concept vector of attribute type according to Algorithm 2, and then process the vector unit according to formula (2). In Steps 4 to 7, under each category, based on formula (4), the similarity simVal of each fuzzy attribute ternary concept vector under this category in VC and FCTA is accumulated in turn, and the number of each category is stored. Step 8 defines the variable maxSim and initialize it to store the maximum

```
Input: network connection $I$ and FCTA
Output: category of $I$
(1)     Preprocessing the network connection $I$
(2) tri$C \longleftarrow \{I, \{w_{ij} | w_{ij} \in I \cap T\}, C_k\}$
(3)     $VC \longleftarrow$ tri$C$
(4)     for each $i \in$ FCTA do
(5)         simValue $\longleftarrow$ Sim( $VC$, $i$ )
(6)         Accumulate the similarity by category
(7)     end for
(8)     maxSim $\longleftarrow 0$, $k \longleftarrow 0$
(9)     for $i = 1$ to $n$ do
(10)        avgSim $\longleftarrow$ Avg (each category simValue)
(11)        if avgSim > maxSim
(12)            maxSim $\longleftarrow$ avgSim
(13)            $k \longleftarrow i$
(14)        end if
(15)    end for
(16)    the classification of $d \longleftarrow k$
```

Algorithm 3: Classification algorithm.

similarity value, defines the variable k, and initializes it to store the condition with the largest similarity, that is, the attack type to which the network connection belongs. Steps 9~15 traverse each value in simVal in turn and calculate its average value and store it in AvgSim. Each value represents the average similarity between VC and the category. We find the value with the largest similarity and store it in MaxSim and store it in MaxSim. The corresponding conditions are stored in C.

The above three algorithms constitute the FCTA, and its time complexity and space complexity are closely related to the number of selected network connections, the number of connection features, and the number of classifications. Assuming that the number of network connections is n, the number of classifications is m, and there are k connection features. When Algorithm 1 initializes text data, the algorithm time complexity is $O(n * k) + O(n * m) + O(n * m * k)$, and the overall time complexity of Algorithm 1 is $O(n * m * k)$. Algorithm 2 constructs a triadic concept based on the background data of the triadic concept. Assuming that the number of triadic concepts is t, the time complexity of Algorithm 2 is $O(n * m * t)$. Similar to Algorithms 1 and 2, according to the pseudocode of Algorithm 3, the time complexity can be calculated to be $O(t)$. Because the number of triadic concepts must be greater than or equal to the number of classifications, the space complexity of the network connection classification algorithm is $O(n * m * t)$. Since Algorithm 1 defines a three-dimensional data set with a triadic concept background, its space complexity is $O(n * m * k)$, while the maximum space required for variables defined in Algorithms 2 and 3 is $O(k * t)$. According to the principle of triadic concept construction, it can be known that the Cartesian product of extension and condition is greater than the number of triadic concepts, so the space complexity of the network connection classification algorithm is $O(n * m * k)$.

4. Experimental Results and Analysis

This section verifies the effectiveness of the proposed network intrusion detection model FCTA through experiments, using the IDS-2018 dataset [34, 35] as training and testing datasets in the experiments and compared with the Support Vector Machine (SVM) classification algorithm [36]. Back Propagation Neural Networks (BP-NNs), and the K-Nearest Neighbor (KNN, K-Nearest Neighbor) classification algorithm [37] on this dataset. Taking the accuracy of the classification results [38], the misjudgment rate of the overall detection results, and the intrusion detection rate [39, 40] as the evaluation criteria, the experimental results verify that the FCTA proposed in this study has high accuracy for the classification of network intrusion data.

Experimental environment: 3.4 GHz CPU and 8 GB memory are used on hardware, and Windows10 operating system and Java experimental platform are used on software.

4.1. Experimental Data.

4.1. Experimental Data. Considering that the research on network intrusion detection should reflect the intrusion characteristics and the freshness of the experimental dataset to the greatest extent, this study adopts the IDS-2018 dataset, which is sponsored by the Communication Security Agency (CSE) and the Canadian Network Security Research (CNSR). A network traffic test dataset established as a collaborative project between the Institutes of Technology (CIC).

4.1.1. IDS-2018 Dataset. The IDS-2018 dataset is a diverse and comprehensive benchmark dataset for intrusion detection generated based on creating user profiles by capturing multiple days of network traffic and system logs for each computer. A total of 15 million pieces of data were sorted out in 10 days and 80 features are extracted from the captured traffic using the network traffic generator

CICFlowMeter-V3. It not only satisfies the experiment needs enough variety to train the detection model as accurate as possible but also covers almost all major network intrusion types, including Benign, Web Attacks, Brute Force, Botnet, DDoS, DoS, infiltration, and SQL injection. These attack types can be broken down into 15 specific attack types. After data analysis and processing by researchers, in the IDS-2018 dataset, the attribute representing time is removed, and each network connection is processed into a data record containing 78 network connection features and 1 data classification.

4.1.2. Data Preprocessing. The experiment is divided into two parts: the first part of the experiment is to compare the FCTA model proposed in this study with the SVM, KNN, and BP-NN algorithms, and the second part of the experiment is to verify the classification effect of the FCTA model under different scales of the IDS-2018 dataset.

Due to the large amount of benign access type data in the original data, more than 12 million, a total of 5 million pieces of benign access data and all attack type data were selected for experimental efficiency. In the first part of the experiment, using the same dataset to train the three algorithms and the FCTA of this study to obtain their respective classification models, one-fifth of the 5 million pieces of data is randomly selected as the training data for the first part of the experiment to train the network intrusion detection model. The distribution of the experimental data is shown in Figure 4. Four groups of 1 million data were selected from the remaining data that did not participate in training as the test dataset, and the average value was calculated to compare the classification effect.

In the second part of the experiment, 5 groups of data, 50,000, 100,000, 200,000, 500,000, and 1 million, were taken from each network intrusion type according to the above proportions to train the model, respectively. The corresponding fuzzy attribute ternary concept vector model FCTA uses the same test dataset for testing; the test set has never participated in training the model data and is selected according to the above ratio. In order to verify the influence of different scales of data on the experimental results, this experiment selected 50,000, 100,000, 200,000, and 500,000 data to form 4 sets of test datasets.

Each piece of data in the IDS-2018 dataset is a network connection, and the value of each network connection feature of the network connection needs to be processed into a continuous numerical type. The traditional method is to use discrete text digital representation and then to convert it into a numerical value between 0 and 1. This method is too subjective and cannot well represent the meaning of the original text. In this study, the network connection feature is processed according to Algorithm 1, which not only converts the discrete text data into numerical data but also obtains the weight value of the network connection feature.

4.2. Evaluation Indicators. Considering the text classification performance evaluation standard and the evaluation standard of the intrusion detection effect, this experiment uses the classification accuracy rate, the misjudgment rate, and the intrusion detection rate as the performance evaluation criteria of the experimental results. The classification accuracy rate represents the rate at which the algorithm can correctly classify the test sample data, which is equal to the ratio of the number of correctly classified samples to the total number of test samples. The calculation formula is shown as

$$\text{Accuracy rate} = \frac{\text{Number of correctly classified samples}}{\text{Total number of samples}} * 100\%. \tag{7}$$

The misjudgment rate is a concept in network intrusion detection in the network field. It represents the ratio of the number of samples of normal types that are judged to be one of the four types of intrusion to the number of normal types of test samples. Its calculation formula is as

$$\text{Misjudgment rate} = \frac{\text{Number of normal samples judged as intrusions}}{\text{Total number of normal type samples}} * 100\%. \tag{8}$$

Intrusion detection rate is also a concept in network intrusion detection. It represents the ratio of the number of correctly classified intrusion samples to the total number of intrusion samples in the test sample. The calculation formula is as follows:

$$\text{Intrusion detection rate} = \frac{\text{Number of correctly classified intrusion samples}}{\text{Total number of invasion type samples}} * 100\%. \tag{9}$$

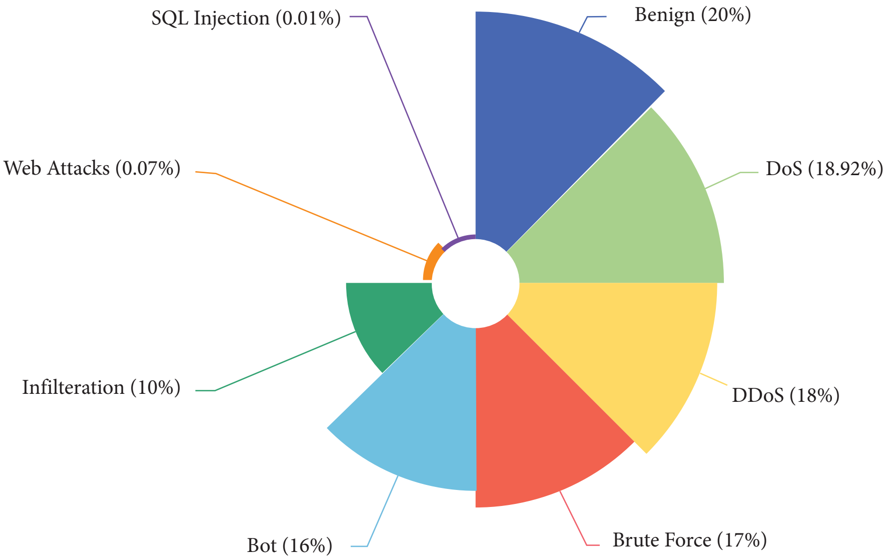

Figure 4: Network intrusion detection training data.

In the above three evaluation criteria, the higher the accuracy rate and the intrusion detection rate, the lower the misjudgment rate and the better the performance of the algorithm. The accuracy rate reflects the correct rate of the algorithm for the overall classification of network connections, and the intrusion detection rate and misjudgment rate reflect the classification effect of the algorithm on the four types of network intrusion data.

4.3. Experimental Results and Analysis

4.3.1. Classification Algorithm Comparison Experiment. Comparing the FCTA proposed in this study with the classic SVM multiclassification algorithm, KNN algorithm, and BP neural network algorithm in machine learning, we use the training dataset to train these four models; then, we use the same test data set to test each model, calculate the accuracy rate, misjudgment rate, and intrusion detection rate of each set of test data, and finally calculate the average value of the three evaluation indicators. Among them, the SVM multiclassification algorithm uses the Linear kernel function, and the rest of the hyperparameters go to the default values; the nearest neighbor k of the KNN algorithm is 5; the BP neural network algorithm uses the Hinge loss function and the stochastic gradient descent algorithm and sets two hidden layers of 64 and 32 neurons, respectively. The results of the comparison experiment's accuracy rate, misjudgment rate, and intrusion detection rate are shown in Figure 5.

Under the same experimental dataset and experimental environment, the average accuracy and average intrusion detection rate of FCTA classification are much higher than BP neural network algorithm, SVM algorithm, and KNN algorithm, reaching 95.41% and 94.75%, respectively, and the FCTA misjudgment rate is 0.97%, which is much lower than the BP neural network algorithm's 12.37% misjudgment rate, the KNN algorithm's 8.57% misjudgment rate, and the SVM algorithm's 2.33% misjudgment rate. Through the above comparative experiments, it is shown that the FCTA proposed in this study has a good classification effect.

Since the IDS-2018 dataset includes eight types of data, Benign, Web Attacks, Brute Force, Botnet, DDoS, DoS, infiltration, and SQL injection. In order to verify the classification accuracy of the FCTA model on each type, five types of benign, Botnet, DDoS, Dos, and infiltration with relatively large amounts of data were selected to compare the FCTA model with the Improved Negative Selection Algorithm (INSA) in [41]. The results of the comparison experiment are shown in Figure 6.

In this comparative experiment, the FCTA proposed in this study has a good classification effect when detecting the above five types of data. When classifying data of Botnet and infiltration attack types, the classification accuracy of FCTA algorithm is slightly lower than that of INSA algorithm, but the overall detection accuracy of FCTA algorithm is higher than that of INSA algorithm. Because the data of Botnet and infiltration types in the training dataset used in the experiment accounted for a relatively low proportion and other types of data accounted for more than 90% in total, the size of the data volume directly affects the FCTA model, so the FCTA model fails to reflect the classification advantage for small sample data. However, the network connection data are easy to collect and the quantity is huge. With the increase of the data volume, the detection accuracy of the FCTA model has a good classification effect. In order to verify the detection effect of FCTA after the scale of the dataset is enlarged, this study conducts the second part of the experiment, that is, the comparison experiment of FCTA classification effect under different dataset scales.

4.3.2. Experimental Results of FCTA Datasets of Different Scales. Under five dataset sizes, we use the intrusion detection models trained by the FCTA algorithm and then perform classification test experiments on the obtained models on the four test sets to calculate the accuracy rate, the

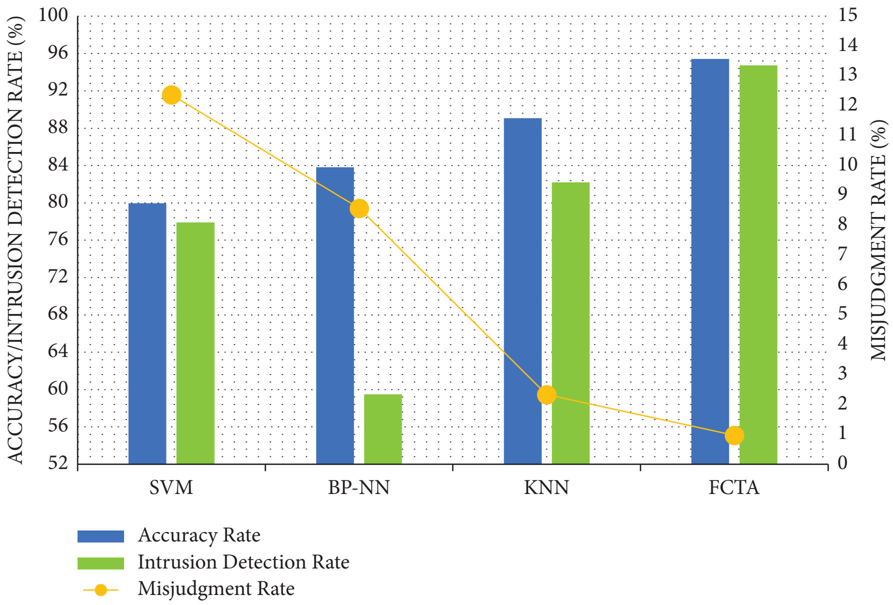

Figure 5: The accuracy rate, intrusion detection rate, and misjudgment rate of classification results' chart.

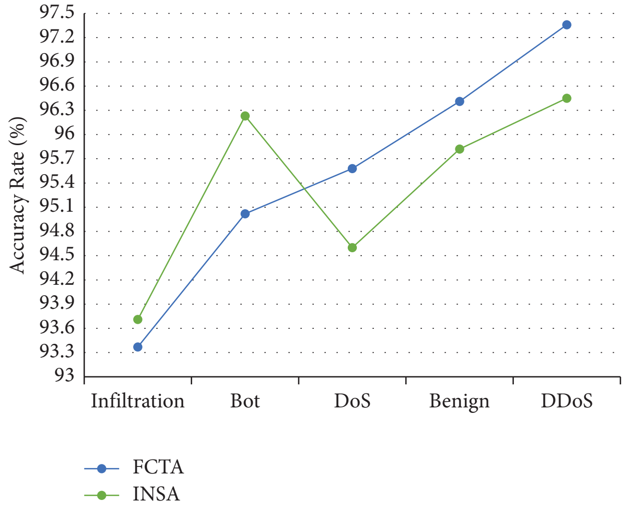

Figure 6: Distribution of accuracy of five types of classification experiments.

Table 2: Experimental results of five datasets of FCTA algorithm.

	Accuracy rate (percent)				Misjudgment rate (percent)				Intrusion detection rate (percent)			
	A	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*
*T*1	84.07	91.13	92.40	93.87	3.75	2.03	3.04	2.19	81.03	89.42	91.26	92.89
*T*2	84.47	93.01	93.50	94.02	2.25	1.73	1.58	1.94	81.15	91.70	92.27	93.01
*T*3	82.80	94.06	93.64	94.31	3.20	1.84	1.40	1.99	79.30	93.04	92.40	93.39
*T*4	81.33	93.70	94.43	94.58	2.70	1.82	1.91	1.70	77.34	92.58	93.51	93.65
*T*5	82.26	93.02	94.61	95.51	2.40	1.98	1.99	0.98	78.43	91.77	93.76	94.63

misjudgment rate of the overall detection result, and the intrusion detection rate. The calculation results are shown in Table 2. *T*1, *T*2, *T*3, *T*4, and *T*5, respectively, represent 50,000, 100,000, 200,000, 500,000, and 1,000,000 datasets. And *A*, *B*, *C*, and *D* represent 10,000, 50,000, 200,000, and 500,000 test datasets.

The comparison charts of the accuracy rate, misjudgment rate, and intrusion detection rate of the experimental results are shown in Figures 7–9, respectively.

It can be clearly seen from the comparison chart of experimental results that the accuracy rate and intrusion detection rate of FCTA are stable within a certain range with

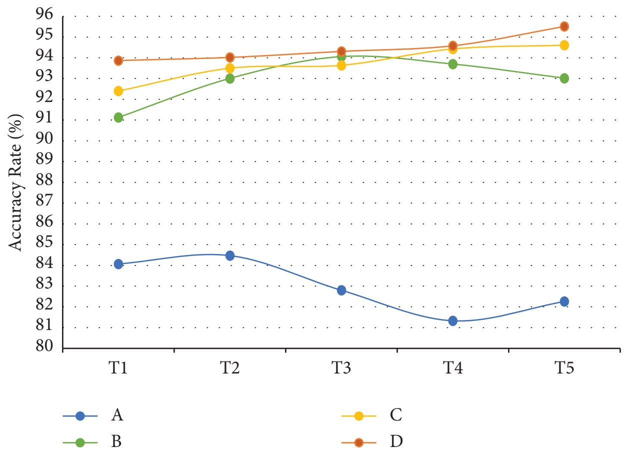

Figure 7: Accuracy rate graph of FCTA.

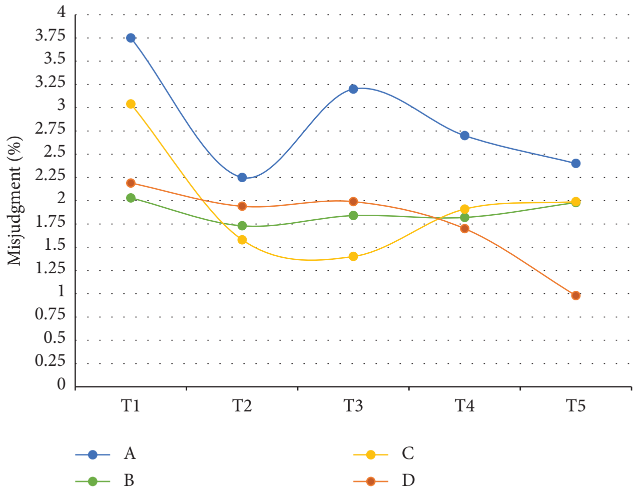

Figure 8: Misjudgment rate graph of FCTA.

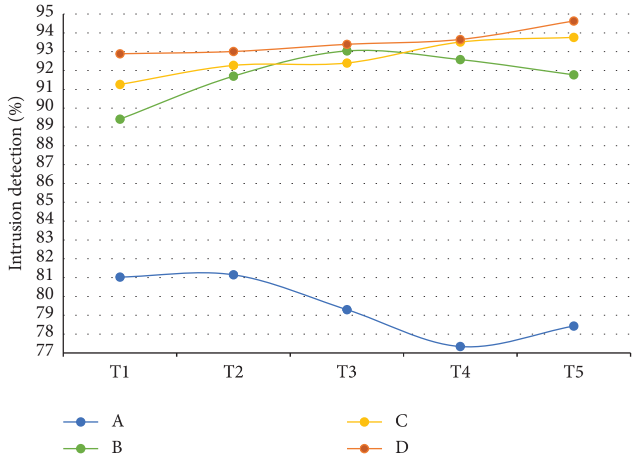

Figure 9: Intrusion detection rate graph of FCTA.

the expansion of the training data size, and the misjudgment rate decreases to less than 2.5% with the increase of the training data amount. Moreover, with the increase of data volume, the accuracy rate and intrusion detection rate increase significantly.

5. Conclusion

For the problem of network intrusion detection, this study cites the triadic concept analysis theory, processes the network connection data into a fuzzy triadic background, and constructs a fuzzy ternary containing the triadic relationship between network connections, network connection characteristics, and intrusion types. Through induction operators concept, we construct the fuzzy attribute ternary concept vector, then use the Euclidean distance formula to calculate the similarity between the fuzzy attribute triadic concept vector and the new sample, and classify the new sample. Finally, in the experiment, this study not only compares the classification effect of FCTA with SVM, KNN, and BP neural network but also verifies that the classification effect of the FCTA model is better under the condition of large samples. Experiments show that the attribute class triadic concept vector model proposed in this study has high accuracy, intrusion detection rate, and low misjudgment rate.

This study uses the triadic concept analysis to resolve the problem of network intrusion detection. The triadic concept analysis has a relatively broad research and improvement space. With the development of technology and the efforts of researchers, the triadic concept analysis will be obtained in the research of network intrusion detection further development.

Data Availability

The study uses the IDS-2018 dataset, which is sponsored by the Communication Security Agency (CSE) and the Canadian Network Security Research. A network traffic test dataset was established as a collaborative project between the Institutes of Technology (CIC). Data are made publicly available at https://www.unb.ca/cic/datasets/ids-2018.html.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by State Key Laboratory of Mathematical Engineering and Advanced Computing.

References

[1] T. Liu, J. Tian, J. Wang et al., “Integrated security threats and defense of cyber-physical systems,” *Acta Automatica Sinica*, vol. 45, no. 1, pp. 5–24, 2019.

[2] Ge Guo, W. Zhang, and B. Zhou, “Preface to the column “theory, method and application of cyber-physical system,” *Control and Decision*, vol. 34, no. 11, pp. 2273–2276, 2019.

[3] K. Scarfone and P. Mell, *Guide to Intrusion Detection and Prevention Systems (IDPS)*, vol. 800, p. 94, NIST Special Publication, 2007.

[4] N. Shone, T. N. Ngoc, and V. D. Phai, “A deep learning approach to network intrusion detection,” *IEEE transactions on emerging topics in computational intelligence*, vol. 2, no. 1, pp. 41–50, 2018.

[5] N. Gao, L. Gao, Y. He, and H Wang, “A lightweight intrusion detection model based on autoencoder network with feature reduction,” *Acta Electronica Sinica*, vol. 45, no. 03, pp. 730–739, 2017.

[6] J. Chen, X. Gao, R. Deng, Y. He, C. Fang, and P. Cheng, “Generating adversarial examples against machine learning based intrusion detector in industrial control systems,” *IEEE Transactions on Dependable and Secure Computing*, vol. 19, p. 1, 2020.

[7] D. Bhamare, M. Zolanvari, A. Erbad, R. Jain, K. Khan, and N. Meskin, “Cybersecurity for Industrial Control Systems: A survey,” *Computers & Security*, p. 89, 2020.

[8] H. Choi, M. Kim, G. Lee, and W. Kim, “Unsupervised learning approach for network intrusion detection system using autoencoders,” *The Journal of Supercomputing*, vol. 75, no. 9, pp. 5597–5621, 2019.

[9] R. U. Khan, X. Zhang, M. Alazab, and R. Kumar, “An Improved Convolutional Neural Network Model for Intrusion Detection in networks,” in *Proceedings of the 2019 Cybersecurity and Cyberforensics Conference (CCC)*, pp. 74–77, IEEE, Melbourne, VIC, Australia, May 2019.

[10] B. Liu, *Research on Intrusion Detection Method of Industrial Control Network Based on Machine Learning*, Harbin Institute of Technology, China, 2020.

[11] W. H. Lin, H. C. Lin, P. Wang, B. H. Wu, and J. Y. Tsai, “Using Convolutional Neural Networks to Network Intrusion Detection for Cyber threats,” in *Proceedings of the 2018 IEEE International Conference on Applied System Invention (ICASI)*, pp. 1107–1110, IEEE, Chiba, Japan, April 2018.

[12] P. Wu and H. Guo, “LuNET: A Deep Neural Network for Network Intrusion detection,” in *Proceedings of the 2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 617–624, IEEE, Xiamen, China, December 2019.

[13] S. Ganapathy, P. Yogesh, and A. Kannan, *Intelligent Agent-Based Intrusion Detection System Using Enhanced Multiclass SVM*, Computational intelligence and neuroscience, vol. 2012, , Article ID 850259, 10 pages, 2012.

[14] S. Ganapathy, P. Vijayakumar, and P. Yogesh, “An intelligent CRF based feature selection for effective intrusion detection,” *The International Arab Journal of Information Technology*, vol. 13, no. 1, 2016.

[15] S. Ganapathy, K. Kulothungan, and S. Muthurajkumar, “Intelligent feature selection and classification techniques for intrusion detection in networks: a survey,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, pp. 1–16, 2013.

[16] B. Riyaz and S. Ganapathy, “A deep learning approach for effective intrusion detection in wireless networks using CNN,” *Soft Computing*, vol. 24, no. 22, pp. 17265–17278, 2020.

[17] F. Lehmann and R. Wille, “A Triadic Approach to Formal Concept analysis,” in *Proceedings of the International Conference on Conceptual Structures*, pp. 32–43, Springer Berlin Heidelberg, Santa Cruz, CA, USA, Augeust 1995.

[18] Li Jinhai, Wei Ling, and Zhang Zhuo, “Concept lattice theory and method and their research prospect,” *Pattern Recognition and Artificial Intelligence*, vol. 33, no. 07, pp. 619–642, 2020.

[19] Z. Li, Z. Zhang, and L. Wang, “Research on text classification algorithm based on triadic concept analysis,” *Computer Science*, vol. 44, no. 8, pp. 207–215, 2017.

[20] W. Ling, Q. Wan, T. Qian, and J. Qi, “An overview of triadic concept analysis,” *Journal of Northwest University*, vol. 44, no. 5, pp. 689–699, 2014.

[21] X.-J. Liu, *Study on the Construction Algorithm of Concept Trilattices and its application*, Xidian University, Xi’an, 2013.

[22] H. Z. Wang and Zhuo Wang, “Liming,” *Application* R*esearch on coalition of triadic concept analysis*, vol. 39, no. 12, pp. 2571–2576, 2018.

[23] X. Li, *Research and Implementation of Intelligent Detection Method of Network Intrusion*, University of Electronic Science and Technology of China, China, 2022.

[24] R. Chapaneri and S. Shah, “A comprehensive survey of machine learning-based network intrusion detection,” in *Proceedings of the Second International Conference on SCI 2018*, Delhi, India, 2019.

[25] X. Li, *Research on Network Intrusion Detection Method Based on Deep Learning*, Lanzhou University of Technology, China, 2020.

[26] C. Xiang, *Research on Network Intrusion Detection Methods under Open Collections*, University of Science and Technology of China, China, 2021.

[27] B. Mukherjee, L. T. Heberlein, and K. N. Levitt, “Network intrusion detection,” *IEEE network*, vol. 8, no. 3, pp. 26–41, 1994.

[28] B. Mukherjee, L. T. Heberlein, and K. Levitt, “Network intrusion detection,” *Network IEEE*, vol. 8, no. 3, pp. 26–41, 1994.

[29] P. Garcia-Teodoro and J. Diaz-Verdejo, “Anomaly-based network intrusion detection: techniques, systems and challenges,” *Computers & Security*, vol. 28, no. 1-2, pp. 18–28, 2009.

[30] D. E. Denning, “An Intrusion-Detection Model,” *an Intrusion-Detection Model*, vol. SE–13, IEEE Computer Society, 1986.

[31] C. H. Huang, J. Yin, and F. Hou, “A text similarity measurement combining word semantic information with TF-IDF method,” *Chinese Journal of Computers*, vol. 34, no. 05, pp. 856–864, 2011.

[32] Osickap, *Concept Analysis of Three-Way Ordinal matrices*, Palacky University, Olomouc, 2012.

[33] C.-K. Li and A. Zaharia, “Induced operators on symmetry classes of tensors,” *Transactions of the American Mathematical Society*, vol. 354, no. 2, 2002.

[34] Cic-Ids2018 and Cse-Cic-Ids2018 on A. W. S., 2021, https://www.unb.ca/cic/datasets/ids-2018.html.

[35] R. M. S. Sigamani and Ganapathi, “P. GOF-SLFN-an intelligent attack detection system against denial of service (DoS) attacks based on glow worm swarm optimized single layer feed forward networks for vehicular cyber physical systems (VCPS),” *IOP Conference Series: Materials Science and Engineering*, vol. 925, no. 1, Article ID 012001, 2020.

[36] T. F. Zhang, Q. F. Fan, and W. Liu, “A support vector machine-based intrusion detection method for SCADA system,” *Journal of Chemical Automation*, vol. 42, no. 02, pp. 153–156, 2015.

[37] J. D. Ren, “An multi-level intrusion detection method based on KNN outlier detection and random forests,” *Journal of Computer Research and Development*, vol. 56, no. 03, pp. 566–575, 2019.

[38] G. H. Feng, “Review of performance evaluation of text classification,” *Journal of Intelligence*, vol. 30, no. 08, pp. 66–70, 2011.

[39] Z. Shi, K. Taghi, and S. Naeem, “Clustering-based network intrusion detection,” *International Journal of Reliability, Quality and Safety Engineering*, vol. 14, no. 2, 2007.

[40] M. Luo, L. Wang, and H. G. Zhang, “An unsupervised clustering-based intrusion detection method,” *Acta Electronica Sinica*, vol. 11, pp. 1713–1716, 2003.

[41] L. Jia, C. Yang, and L. L. Song, “Improved negative selection algorithm and its application in intrusion detection,” *Computer Science*, vol. 48, no. 06, pp. 324–331, 2021.