IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Automatic classification method for software vulnerability based on deep neural network

Guoyan Huang[1], Yazhou Li[1], Qian Wang[1,*], Jiadong Ren[1], Yongqiang Cheng[3], Xiaolin Zhao[2]

[1]Computer Virtual Technology and System Integration Laboratory of Hebei Province, College of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei, China, 066000
[2]Beijing Key Laboratory of Software Security Engineering Technology, School of computer Science and Technology, Beijing Institute of Technology, Beijing, China, 100081
[3]Computer Science, University of Hull, Hull, UK, HU6 7RX

Corresponding author: Qian Wang (e-mail: wangqianysu@163.com).

This work is supported by the National Key R&D Program of China under Grant No.2016YFB0800700, the National Natural Science Foundation of China under Grant Nos.61472341, 61772449, 61572420, 61807028 and 61802332, the Natural Science Foundation of Hebei Province China under Grant No.F2016203330. The authors are grateful to valuable comments and suggestions of the reviewers.

ABSTRACT Software vulnerabilities are the root causes of various security risks. Once a vulnerability is exploited by malicious attacks, it will greatly compromise the safety of the system, and may even cause catastrophic losses. Hence automatic classification methods are desirable to effectively manage the vulnerability in software, improve the security performance of the system and reduce the risk of the system being attacked and damaged. In this paper, a new automatic vulnerability classification model (TFI-DNN) has been proposed. The model is built upon term frequency- inverse document frequency (TF-IDF), information gain (IG) and deep neural network (DNN): the TF-IDF is used to calculate the frequency and weight of each word from vulnerability description; the IG is used for feature selection to obtain an optimal set of feature word; and the DNN neural network model is used to construct an automatic vulnerability classifier to achieve effective vulnerability classification. The National Vulnerability Database (NVD) of the United States has been used to validate the effectiveness of the proposed model. Compared to SVM, Naive Bayes and KNN, the TFI-DNN model has achieved better performance in multi-dimensional evaluation indexes including accuracy, recall rate, precision and F1-score.

INDEX TERMS Deep neural network, information gain, software security, vulnerability classification

I. INTRODUCTION

With the rapid development of information technology, the impacts brought to industries by application of the Internet and computers are twofold. They bring convenience but also huge risks and hidden dangers. Recently, with the improvement of the digitalization level of various industries, information security issues have become increasingly prominent. Vulnerabilities are defined as software and hardware defects of the system being illegally exploitable by unauthorized personnel. Once the vulnerability of information system is exploited by malicious attack, the security of information system will be at great risk and may cause inestimable consequences. For example, in 2017, hackers exploited Windows system vulnerabilities to expose 100,000 organizations worldwide to Bitcoin ransomware. In the same year, Microsoft released a total of 372 Office vulnerability patches. Hackers can use the office vulnerabilities to conduct Advanced Persistent Threat (APT) attacks, spread botnets, ransomware and so on. In recent years, the number and variety of vulnerabilities have gradually increased, so the management and analysis of software vulnerabilities has become more and more important.

If the vulnerability can be classified and managed effectively, it can not only improve the efficiency of vulnerability recovery and management, but also reduce the risk of system being attacked and damaged, which is vitally important for the security performance of the system. As software security vulnerabilities play an important role in cyber security attacks, more and more researches on vulnerability classification are conducted by relevant security researchers. The earlier vulnerability classification method RISOS [1], which is aimed at the operating system of the computer, mainly divides the operating system vulnerabilities into seven categories from the perspective of attack, and describes how to exploit the vulnerabilities instead of triggering these vulnerabilities' condition. The PA vulnerability classification method in [2] not only studied the vulnerabilities of the operating system, but also classified the vulnerabilities existing in the application. Andy Gray vulnerability classification method [3] proposed a vulnerability classification system consisting of ten categories according to the different analysis needs of the vulnerability. However, as the complexity of vulnerabilities increases, the limitations of traditional artificial vulnerability classification methods become more and more obvious. Therefore, researchers pay more attention to automatic classification of vulnerabilities.

A large number of machine learning methods have been recently reported in the field of text classification [4]. Classifying them by vulnerability description is also a kind of text classification. Therefore, the problem of automatic classification of vulnerabilities can also be solved using machine learning methods**.** Davari M et al. proposed an automatic vulnerability classification framework based on conditions that activate vulnerabilities [5], different machine learning techniques (Random Forest, C4.5 Decision Tree, Logistic Regression, and Naive Bayes) are employed to construct a classifier with the highest F-measure. The 580 software security flaws of the Firefox project were analyzed through experiments to evaluate the effectiveness of the classification. The SVM classification method based on LDA model [6] is applied in the domain of vulnerability classification by Bo Shuai et al. The SVM based on the topic model can make full use of the number of distributed vulnerabilities for classification. The experiment results indicated that SVM has achieved good results in vulnerability classification. Dumidu Wijayasekara et al. tested the Naïve Bayes classifier by using textual information from the error description [7]. The analysis illustrates the feasibility of the Naïve Bayes classifier to classify textual information based on the vulnerability description. Sarang Na et al. proposed a classification method for classifying CVE entries that could not provide enough information into vulnerability categories using the Naïve Bayes classifier [8]. Marian Gawron et al. applied Naive Bayes algorithm and simplify artificial neural network algorithm to vulnerability classification [9], and made comparison on the same data set. The experimental results showed that the artificial neural network algorithm was superior to Naive Bayes algorithm in vulnerability classification. A data-driven approach to machine learning for vulnerability detection was proposed by Harer J A et al. [10]. They also compare the application of deep neural network models with more traditional models such as random forests, and find that the best performance comes from combining the features of deep model learning with tree-based models. Finally, the highest performance model proposed in the paper achieves an area under the precision-recall curve of 0.49 and an area under the ROC curve of 0.87. It can be seen that a large number of machine learning algorithms have been well applied in the field of vulnerability classification.

Although these machine learning classification algorithms have achieved promising results in many fields, due to the large amount of vulnerability data and short description, the generated word vector space presents the characteristics of high dimension and sparse. These machine learning algorithms are not very effective in dealing with high and sparse problems. At the same time, they ignore specific vulnerability information and the classification accuracy is not high. However, in recent years, deep learning has been applied in many fields and has achieved success [11], such as the field of speech and image recognition [12,13], the error rate in speech recognition is reduced by 20%-30% [14,15], and the error rate in the ImageNet evaluation task is reduced by 26%-15% [16]. For face image classification based on GoogLeNet network [17], the classification of age and gender is achieved with high precision. Deep learning also has a significant impact in the field of natural language [18,19]. Hwiyeol et al. studied the classification problem in the field of natural language, and applied convolutional neural networks (CNN) and recurrent neural networks (RNN) to the field of large-scale text classification and achieved success [20]. Aziguli W et al. proposed a novel text classifier using DNN model [21] to improve the computational performance of processing large text data with mixed outliers. Wu F et al. proposed a deep learning method for vulnerability detection [22], namely, convolution neural network (CNN), long short term memory (LSTM) and convolution neural network-long short term memory (CNN-LSTM), and the vulnerability prediction accuracy reached 83.6%, which is superior to the traditional method. Li Z et al. designed and implemented a deep learning-based vulnerability detection system [23], which alleviated the cumbersome and subjective tasks of human experts manually defining features. Therefore, deep learning can also be applied to the field of software vulnerability detection and achieved good results.

Therefore, in order to better deal with the high and sparse word vector space and take advantage of the automatic feature extraction by deep learning, this paper proposes an automatic vulnerability classification model TFI-DNN based on term frequency-reverse document frequency (TF-IDF), information gain (IG) and deep neural network (DNN). In the model, we first use TF-IDF-IG (TFI) algorithm to extract the feature of the description text and reduce the dimension of the generated high-dimensional word vector space, then construct a DNN neural network model based on deep learning. The TFI-DNN model was trained and tested using vulnerability data from the National Vulnerability Database (NVD). The test results show that the automatic vulnerability classification model in this paper effectively improves the performance of vulnerability classification.

The rest of this paper is arranged as follows. Section 2 introduces the definition of relevant algorithms. In section 3, the implementation details of our model are described. Experiment dataset and results are discussed in Section 4 with comparative analysis. Section 5 outlines the conclusions of this paper and potential future research.

II. RELATED DEFINITION

The automatic classification model of vulnerability (TFI-DNN) is constructed in this paper. The relevant definitions are as follows.

A. TF – IDF

TF-IDF is a common weighted technology based on statistical methods [24]. Assume there are a set of files and each file contains a number of words. We define the importance of the word i in file j as follows.

$$\mathrm{tf}_{\mathrm{ij}} = \frac{\mathrm{n}_{\mathrm{i,j}}}{\sum_{\mathrm{k}} \mathrm{n}_{\mathrm{k,j}}} \tag{1}$$

Where both i and j are positive integers, $\mathrm{n}_{\mathrm{i,j}}$ denotes the frequency of word i in the file j

The formula for IDF is as follows.

$$\mathrm{idf}_{\mathrm{i}} = \log\frac{|\mathrm{F}|}{|\{\mathrm{j}:\mathrm{t}_{\mathrm{i}} \in \mathrm{d}_{\mathrm{j}}\}|} \tag{2}$$

Where $|\mathrm{F}|$ represents the total number of files in the corpus, f_{j} is the jth file, and $|\{\mathrm{j}: \mathrm{t}_{\mathrm{i}} \in \mathrm{f}_{\mathrm{j}}\}|$ represents the number of files containing the word t_{i} .

The formula for TF-IDF is as follows.

$$\mathrm{TF} - \mathrm{IDF} = \mathrm{tf}_{\mathrm{ij}} * \mathrm{idf}_{\mathrm{i}} \tag{3}$$

TF-IDF is used to evaluate the importance of a word to a document in the document set or in a corpus. The importance of a word increases proportionally with the number of times it appears in the file, but it also decreases inversely with the frequency it appears in the corpus.

B. INFORMATION GAIN (IG)

IG [25] refers to that, if a feature X in class Y is known, the information uncertainty of class Y will decrease, and the reduced uncertainty degree reflects the importance of feature X to class Y. Set the training data set to D, $|\mathrm{D}|$ indicates the number of samples in D. Suppose there are K classes C_{k}, k =1, 2,…, K， $|\mathrm{C}_{\mathrm{k}}|$ is the number of samples belonging to class C_{k} ， $\sum_{\mathrm{C}_{\mathrm{k}}=1}^{\mathrm{K}}|\mathrm{C}_{\mathrm{k}}| = |\mathrm{D}|$. If feature A has n different values $\{\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_{\mathrm{n}}\}$, D will be divided into n subsets according to the values in feature A, denoted as D= ($\mathrm{D}_1, \mathrm{D}_2, \ldots,$ D_{n}), where $|\mathrm{D}_{\mathrm{i}}|$ is the number of samples in D_{i}, $\sum_{\mathrm{i}=1}^{\mathrm{n}} |\mathrm{D}_{\mathrm{i}}| = |\mathrm{D}|$. The set of samples belonging to class C_{k} in D_{i} is D_{ik} , $\mathrm{D}_{\mathrm{ik}} = \mathrm{D}_{\mathrm{i}} \cap \mathrm{D}_{\mathrm{k}}$, $|\mathrm{D}_{\mathrm{ik}}|$ is the number of samples of D_{ik}.

The empirical entropy H (D) of data set D is calculated as follows.

$$H(D) = -\sum_{k=1}^{K} \frac{|C_k|}{|D|} \tag{4}$$

The empirical conditional entropy H (D|A) of feature A for dataset D is calculated as follows.

$$H(D|A) = -\sum_{i=1}^{n} \frac{|D_i|}{|D|} \sum_{k=1}^{K} \frac{|D_{ik}|}{|D_i|} \log_2 \frac{|D_{ik}|}{|D_i|} \tag{5}$$

The information gain calculation formula for each feature is as follows.

$$g(D,A) = H(D) - H(D|A) \tag{6}$$

According to the feature selection method of the information gain criterion, the information gain of each feature is calculated, and the features with larger information gain value are selected.

C. FORWARD PROPAGATION ALGORITHM

The forward propagation algorithm uses a number of weight coefficient matrices W, the offset vector b and the input value vector x to perform a series of linear operations and activation operations, starting from the input layer to the output layer to get the final output. The calculation formula is as follows.

$$x^l = \sigma(z^l) = \sigma(W^l x^{l-1} + b^l \tag{7}$$

Where W is the matrix corresponding to the hidden layer and the output layer, b^{l} is the offset vector, x is the input value vector, l is the number of layers of the neural network, and x^{l} is the output result.

D. ACTIVATION FUNCTION

softmax and tanh are commonly used activation functions in multiple classification problems，they can map the output of multiple neurons to the range of [0,1]. The calculation formulas are as follows.

$$\mathrm{softmax}(\mathrm{y}_{\mathrm{i}}) = \frac{\mathrm{e}^{\mathrm{y}_{\mathrm{i}}}}{\sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{e}^{\mathrm{y}_{\mathrm{j}}}} \tag{8}$$

$$\tanh(y_i) = \frac{1-e^{-2y_i}}{1+e^{-2y_i}} \tag{9}$$

Where, y_{i} represents output of the neural network.

E. CROSS ENTROPY COST FUNCTION

Cross entropy describes the distance between two probability distributions, i.e. the distance between the actual output and the expected output. The calculation formula is as follows.

$$L = -\frac{1}{n}\sum_n |y\ln(a) + (1-y)\ln(1-a)| \tag{10}$$

Where, y is the desired output, a is the actual output of the neuron and n is the number of categories.

The smaller the value of the cross entropy, the closer the two probability distributions are. It is a widely used loss function in the classification problem.

F. BATCH GRADIENT DESCENT ALGORITHM (BGD)

The gradient descent algorithm is used on all the parameters to make the loss function reach a smaller value, and finally all the parameters of the model with the minimum loss are obtained. The calculation formula is as follows.

$$\theta_i := \theta_i - \alpha \sum_{j=0}^{m} (h_\theta\left(x_0^{(j)}, x_1^{(j)}, \ldots x_n^{(j)}\right) - y_j)\, x_i^{(j)} \tag{11}$$

Where α and θ_{i} represent the learning rate and the i th parameter, θ represents all parameters, m is the sample number, y_{j} is the actual category and $\mathrm{h}_\theta\left(\mathrm{x}_0^{(\mathrm{j})}, \mathrm{x}_1^{(\mathrm{j})}, \ldots \mathrm{x}_{\mathrm{n}}^{(\mathrm{j})}\right)$ is the output of sample j on θ. $\mathrm{x}_{\mathrm{i}}^{(\mathrm{j})}$ is ith feature of sample j.

III. THE TFI-DNN ALGORITHM

The vulnerability automatic classification model TFI-DNN is mainly composed of TFI and DNN. The original vulnerability data is preprocessed first, then TFI is used to extract the features of the vulnerability description text and reduce the dimensionality of the generated higher-dimensional word vector space, and then the DNN is constructed to realize the automatic training and classification of the vulnerability. TFI-DNN algorithm is shown in Figure 1.

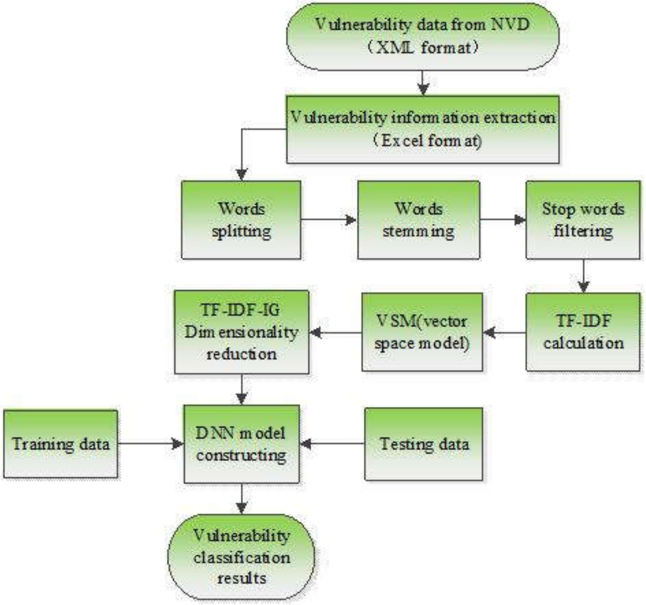

FIGURE 1. **Framework of the Method.**

A. FEATURE SELECTION USING TFI

TFI is mainly used to extract feature word set. The steps are as follows.

Input: word list (word_list) formed by word segmentation, lemmatiation, and stop word filtering.
Output: feature word set (feature_words).

1) Traversing each word in the word_list.
2)Word frequency statistics for word_list, stored in the doc_frequency list.
3) Traversing the word frequency list doc_frequency.
4) Calculate the TF value of each word according to (1) and store it in the word_tf dictionary.
5) Calculate the IDF value of each word according to (2) and store it in the word_idf dictionary.
6) Calculate the TF-IDF value of each word according to (3) and store it in the word_tf_idf dictionary.
7) The word set is sorted in descending according to the TF-IDF value.
8) Select the first n words as an important feature set.
9) Save important words in the feature list (features_vocabSet).
10) Traverse features_vocabSet, divide features_vocabSet and store the subset into the subDataSet.
11) Calculate probability of subDataSet.
12) Calculate the empirical conditional entropy of each word according to (4) and (5) and store it in newEntropy.
13) Calculate the IG value of each word according to (6).
14) Save each word and the corresponding IG value in the dictionary.
15) The word set is sorted descending by IG value.
16) Select the first m words as features and store them in the feature_words.
17) Return feature_words.

B. OPTIMIZATIONS USING DNN

DNN consists of one input layer, multiple hidden layers and one output layer, whose input is the feature vector of instance and output is the category of instance. It mainly includes two propagation processes, forward propagation and back propagation. The propagation process is as follows.

Input: training set after vectorization.
Output: coefficient matrix W and offset vector b of DNN.

1) Input training data set x, initialize weight coefficient matrix W and offset vector b.
2) Perform a linear calculation for each neuron xi, the calculation formula is $z = \sum_{i=1}^{m} W_i x_i + b$, use tanh for the activation operation, then calculate the final output result layer by layer according to (7). The result output is $a^l = \sigma(z^l) = \sigma(W^l a^{l-1} + b^l)$.
3) Softmax is used to assign probabilities to different objects for the final output, and the output of multiple neurons is mapped to the range of [0,1] .
4) The cross entropy is selected as the loss function to measure the loss between the output calculated by the training sample and the actual training sample output.
5) Select the batch gradient descent algorithm as the optimization algorithm, minimize the loss function, set the batch size, iteration threshold and learning rate $\alpha(0 < \alpha \leq 1)$, the DNN model was trained iteratively.
6) Finally, output optimized DNN model parameters W and b.

IV. EXPERIMENTAL ENVIRONMENT AND DATA SET

A. ENVIRONMENT

The experiment was conducted on PC with Intel(R) Core(TM) i5-4460 processor, 3.20 GHz and 8.00 GB memory, running Windows 7 operating system. Programming uses Pycharm 2017 on Anaconda3 version and the TensorFlow framework.

B. DATA SET OF EXPERIMENT

In order to verify the effectiveness, we use the internationally recognized National Vulnerability Database (NVD) [26] as experimental data. The source file of this dataset is a series of XML files, which contains comprehensive information about the vulnerability, such as CVE number, vulnerability release date, CVSS_version, CVSS_score, CVSS_vector, vulnerability text description, and so on. The annual vulnerability amount (2000-2016) of NVD vulnerability database is shown in Figure 2, and the total number of vulnerability records is 56,838.

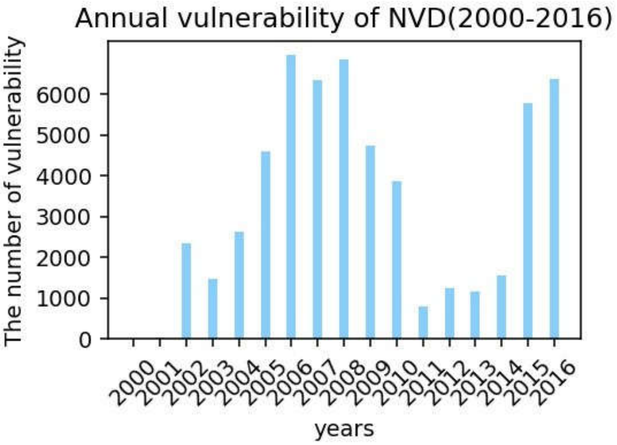

FIGURE 2. Annual Vulnerability of NVD (2000-2016).

By the end of 2016, the NVD vulnerability database contained 23 vulnerability categories and 1 “unknown” vulnerability category. The number of vulnerabilities records in each category is highly variable. For example, there are 11284 vulnerabilities records in the "unknown" category whist there are only 5 vulnerabilities records in the "environment conditions" category. As the "unknown" category of vulnerability has no value for the model training, the vulnerability in "unknown" category is excluded from the experiment. The other 23 vulnerability categories with clearly identified categories are included. The distribution of vulnerability records in each category is shown in Figure 3.

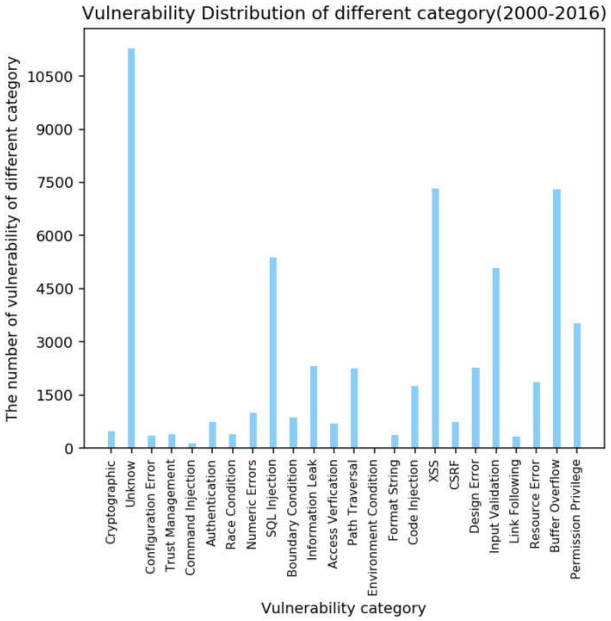

FIGURE 3. Vulnerability Distribution of different category.

The required vulnerability information are extracted from the XML vulnerability files using program codes written in Python, including CVE number, vulnerability text description, and vulnerability category. All the text information of vulnerabilities from 2000 to 2016 is collected for statistics, 10,000 of them were selected as the training set, and 1,000 of them were used as the test set. Samples of the data are shown in Table I.

TABLE I
EXAMPLES OF DATASETS

CVE_ID	Description	Vulnerability category
CVE-2016-9997	SPIP 3.1.x suffers from a Reflected Cross Site Scripting Vulnerability in involving the ···	XSS
CVE-2016-9993	IBM Kenexa LCMS Premier on Cloud 9.0, and 10.0.0 is vulnerable to SQL injection. A remote attacker could send specially-crafted SQL statements, which could allow the attacker to view, add, modify or delete information in the back-end database. IBM Reference #: 1992067.	SQL Injection
···	···	···

V. EXPERIMENTAL PROCEDURE

A. DATA PREPROCESSING

1) WORD SEGMENTATION

Word segmentation refers to the cutting of coherent vulnerability text information into one word, which transforms the entire vulnerability text information into the smallest semantic unit that can be counted by statistics. This is the first and most important step in the process of vulnerability text preprocessing. For the text of the vulnerability described in English, the word segmentation is very simple. You only need to identify the entire vulnerability description by dividing the space or punctuation between the texts.

For example, the vulnerability text numbered CVE-2016-9990 is described as: "IBM iNotes 8.5 and 9.0 is vulnerable to cross-site scripting. This vulnerability allows users ···". After the word segmentation, you can get the word set as ["IBM", "iNotes", "8.5", "and", "9.0", "is", "vulnerable", "to", "cross", "-" , "site", "scripting", ".", "This", "vulnerability", "allows", "users",···].

2) LEMMATIZATION

Lemmatization refers to the transformation of non-root form into root form in the word set, and that is the verb in English description changing according to the person into the verb prototype. Convert the plural form of a noun to the singular form, convert gerund form to verb prototype, etc. From the perspective of data mining, these words should belong to the same category of semantically similar words. For example, "attack", "attacking", and "attacked" can be divided into the same word, which can be expressed as "attack" in the root form.

Therefore, the set of words in the vulnerability text numbered CVE-2016-9990 is converted by lemmatization: ["IBM", "iNote", "8.5", "and", "9.0", "is", "vulnerable", "to", "cross", "-" , "site", "scripting", ".", "This", "vulnerability", "allow", "user",···]. It can be seen that the following words in the original description information, "iNotes", "allows", "users" are restored to "iNote", "allow", "user", etc.

3) STOP WORD FILTERING

Stop word filtering refers to words that appear frequently in text and contribute little or no contribution to the content or classification of text information, it includes both public stop words and professional stop words. We can download public stop word list from the Internet[27], such as common prepositions, articles, auxiliary words, modal verbs, pronouns, and conjunctions, etc. In addition, some words such as "information", "security", and "vulnerability" are meaningless to the classification of vulnerability， so these words should be filtered out. In this paper, each word of the vulnerability text is arranged in descending order by word frequency, and we set a threshold. Based on our experience, we select words higher than the threshold value and meaningless for classification of vulnerabilities as stop word and add them into the stop word list to form a professional stop word list.

B. FEATURE EXTRACTION

The candidate feature set composed of 17,935 words was obtained from all experimental data through the above data preprocessing operation. First, the weight of each word in the vulnerability candidate feature set is calculated by TF-IDF, and the results are arranged in descending order. The number of feature words of the extracted feature subset is 1024. Part of the arrangement is shown in Table II.

TABLE II
FEATURE WORDS BASED ON TF-IDF

Feature	overflow	SQL	XSS	XSS	traversal	⋯
TF-IDF	0.0215	0.0206	0.0198	0.0110	0.0109	⋯

According to the 1024 words selected in 1), IG is used to calculate the information gain value of each word in the word set, and the results are arranged in descending order. The number of feature words extracted from the feature subset is 900. Part of the arrangement is shown in Table III.

TABLE III
FEATURE WORDS BASED ON IG

Feature	overflow	traversal	SQL	XSS	buffer	⋯
IG	0.6110	0.5168	0.4572	0.4225	0.3039	⋯

C. THE WORD VECTOR SPACE ESTABLISHMENT

In the application of this paper, each vulnerability description is expressed as an m-dimensional vector (m is the number of feature words in the feature word set) [28]. After all the vulnerability text samples are vectorized through the m-dimensional space, the vector representation of the vulnerability is established. Each vulnerability sample can be regarded as a point in the high-dimensional space, where the representation of one vulnerability sample $\mathrm{v_i}$ is $\mathrm{v_i} = \{\mathrm{t_1},\ \mathrm{t_2},\ \ldots,\ \mathrm{t_n}\}$. Among this formula, m is the number of feature words, which is consistent with the dimension of the vector space of vulnerability.

D. DNN MODEL CONSTRUCTION

After the vectorization representation of vulnerability description, the vulnerability text data described by natural language is transformed into the data structure that can be recognized by the machine and expressed through statistical learning. We use the deep learning framework TensorFlow to construct the DNN model. Various configuration parameters of the DNN model were tested, including the number of hidden layers, the number of neurons in each layer, the learning rate of the neural network and the number of iterations. Finally, the parameters are set as empirical values from experiment. The DNN model is shown in Figure 4.

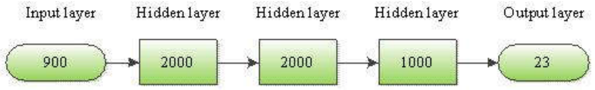

FIGURE 4. The Model of DNN.

The parameters of the DNN model are set as follows.
1) Initialize the weight of DNN model with normal distribution function with standard deviation of 0.1.
2) The offset values of the input layer and hidden layer are set to 0.1.
3) The forward propagation of the input layer and hidden layer uses *tanh* as the activation function, while the output layer uses *softmax* as the activation function.
4) Use the dropout function in TensorFlow to prevent overfitting.
5) The cross entropy function is used as the loss function to measure the loss between the calculated output of the training sample and the actual output of the training sample.
6) Optimize the DNN model using a batch gradient descent algorithm with a batch size set to 100 and a learning rate set to 0.2.
7) The number of iterations of DNN model training is 20.
Next, the training sample of the NVD data set is used to train the TFI-DNN vulnerability automatic classification model, and then the vulnerability test set is used to evaluate the model performance.

VI. RESULT AND COMPARISONS

A. EVALUATION INDEX

In order to evaluate the performance of automatic classification model (TFI-DNN) in this paper and to compare the performance between different algorithms, unified evaluation criteria are needed. According to the general method of data mining theory, the most widely used classification method for multi-class problem evaluation model is multi-class confusion matrix. As shown in Table IV.

TABLE IV
MULTI-CLASS CONFUSION MATRIX

True category	Forecast category		
	Buffer Overflow	Configuration Errors	⋯
Buffer Overflow	True Positive(TP_i)	False Negative(FP_i)	⋯
Configuration Errors	False Positive(FN_i)	True Negative(TN_i)	⋯
⋯	⋯	⋯	⋯

Where i represents a category of a vulnerability. According to the confusion matrix, the accuracy, recall rate, precision rate and F1-score are used as evaluation indexes of TFI-DNN model. The calculation formulas of each index are as follows.

Accuracy $(\boldsymbol{i})$ indicates the ratio of the number of test instances correctly classified by vulnerability i to the total number of test instances. The calculation formula is below.

$$\text{Accuracy}(i) = \frac{TP_i + TN_i}{TP_i + FP_i + FN_i + TN_i}$$

Recall$(\boldsymbol{i})$ indicates the ratio of the number of the positive examples correctly classified by vulnerability i to the number of actual positive examples. The calculation formula is below.

$$\mathrm{Recall(i)} = \frac{\mathrm{TP_i}}{\mathrm{TP_i + FN_i}}$$

Precision$(\boldsymbol{i})$ represents the ratio of the number of positive examples correctly classified by vulnerability i to the number of instances classified as positive examples. The calculation formula is below.

$$\mathrm{Precision(i)} = \frac{\mathrm{TP_i}}{\mathrm{TP_i + FP_i}}$$

F1-score$(\boldsymbol{i})$ combines the recall rate and the precision rate. The calculation formula is below.

$$\mathrm{F1 - score(i)} = \frac{2 * \mathrm{Recall(i)} * \mathrm{Precision(i)}}{\mathrm{Recall(i)} + \mathrm{Precision(i)}}$$

The value of F1-score is between 0 and 1, and the higher the value, the better the performance of the vulnerability classification model.

B. COMPARISONS BETWEEM TF-DIF-DNN AND TFI-DNN

In order to evaluate the effect of TFI for feature word selection, we compare it with the method of TF-IDF without IG, the same configured DNN model was iterated 20 times. The evaluation results of the vulnerability classification in the multidimensional evaluation index are shown in Figure 5.

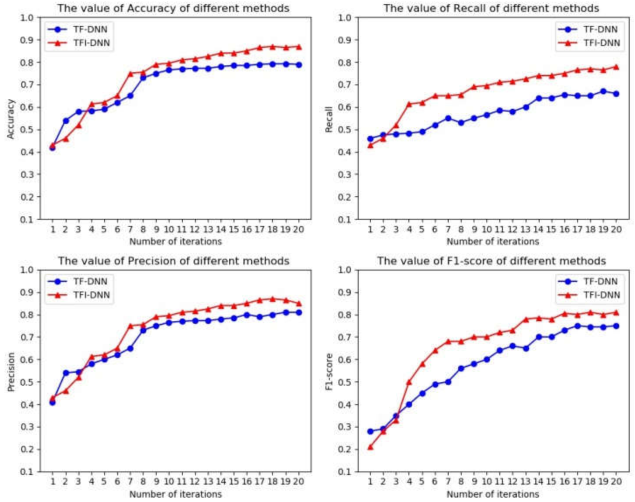

FIGURE 5. **The Value of Evaluation Indexes by Different Methods.**

Figure 5 shows that the DNN model based on TFI at the same number of iterations is superior to the DNN model based on TF-IDF in accuracy, recall rate, precision and F1-socre. Because TF-IDF is not comprehensive and it just considers the word frequency to measure the importance of words, and sometimes important words may not appear many times. TFI not only considers the importance of word frequency, but also takes into account the importance of words to the category of vulnerability, so that a better word set can be further extracted according to the IG value, which makes better performance in various evaluation indexes.

B. COMPARISONS AMONG CLASSIFICATION METHORDS

Many machine learning algorithms, such as SVM[29], Naive Bayes(NB)[10], KNN[30] and so on have been widely used in the field of text classification. However, due to the large amount of vulnerability data and short description, the generated word vector space presents the characteristics of high dimension and sparse. But these machine learning algorithms are not very effective in dealing with high and sparse problems. Therefore, this paper proposes a vulnerability automatic classification model TFI-DNN based on deep learning. In order to evaluate the performance of TFI-DNN in vulnerability classification, we compare and evaluate it with the traditional machine learning algorithm based on TFI in terms of accuracy, recall, precision, and F1-score. The experimental results are shown in Table V.

TABLE V
RESULTS OF DIFFERENT CLASSIFICATION METHODS

Model	Accuracy	Recall	Precision	F1-score
TFI-SVM	0.68	0.68	0.61	0.63
TFI-NB	0.76	0.78	0.72	0.75
TFI-KNN	0.81	0.78	0.76	0.77
TFI-DNN	**0.87**	**0.82**	**0.85**	**0.81**

According to the experimental comparison results, we can see that the TFI-DNN vulnerability automatic classification model achieved an overall better performance than the other three algorithms. Specifically TFI-DNN is 19%, 11% and 6% higher in classification accuracy, and 18%, 6% and 4% higher in F1-score comprehensive evaluation indexes compared to SVM, Naive Bayes and KNN respectively. The experiment results above have validated the effectiveness of the TFI-DNN model in the automatic classification of vulnerabilities.

VII. CONCLUSION

In order to better analyze and manage vulnerabilities according to their belonging classes, improve the security performance of the system, and reduce the risk of the system being attacked and damaged, this paper applied deep neural network to software vulnerability classification. The analysis of the method and construction process of TFI and DNN are discussed in detail. We compared the vulnerability classification model TFI-DNN to TFI-SVM, TFI-Naïve Bayes and TFI-KNN on the NVD dataset. The results show that the proposed TFI-DNN model outperforms others in accuracy, precision and F1-score and performs well in recall rate. And it is superior to SVM, Naïve Bayes and KNN on comprehensive evaluation indexes. The work in this paper shows the effectiveness of TFI-DNN in vulnerability classification, and provides a basis for our future research using the benchmark vulnerability dataset.

ACKNOWLEDGEMENT

This work is supported by the National Key R&D Program of China under Grant No.2016YFB0800700, the National Natural Science Foundation of China under Grant

Nos.61472341, 61772449, 61572420, 61807028 and 61802332, the Natural Science Foundation of Hebei Province China under Grant No.F2016203330. The authors are grateful to valuable comments and suggestions of the reviewers.

REFERENCES

[1] Abbott R, Chin J, Donnelley J, et al, "Security Analysis and Enhancements of Computer Operating Systems." Washington DC, USA: US Department of Commerce, 1976.

[2] Bisbey II R, Hollingworth D, "Protection Analysis: Final Report." Marina Del Rey, USA: University of Southern California, 1978.

[3] A. Gray. An historical perspective of software vulnerability manageme nt.Information Security Technical Report, 2003.8(4):34-44.

[4] Pan Jun Kim, "An Analytical Study on Automatic Classification of Domestic Journal articles Based on Machine Learning." Journal of the Korean Society for Information Management, vol.35, 2018(2): 37-62.

[5] Davari, Maryam , M. Zulkernine , and F. Jaafar . "An Automatic Software Vulnerability Classification Framework." 2017 International Conference on Software Security and Assurance (ICSSA) IEEE Computer Society, 2017: 44-49.

[6] Shuai, Bo, et al. "Automatic classification for vulnerability based on machine learning." IEEE International Conference on Information and Automation IEEE, 2014:312-318.

[7] Wijayasekara, Dumidu, M. Manic, and M. Mcqueen. "Vulnerability identification and classification via text mining bug databases." Industrial Electronics Society, IECON 2014 -, Conference of the IEEE, 2015:3612-3618.

[8] Na, Sarang, T. Kim, and H. Kim. "A Study on the Classification of Common Vulnerabilities and Exposures using Naïve Bayes." Advances on Broad-Band Wireless Computing, Communication and Applications. Springer International Publishing, 2016:657-662.

[9] Gawron, Marian, C. Feng, and C. Meinel. "Automatic Vulnerability Classification Using Machine Learning." International Conference on Risks and Security of Internet and Systems Springer, Cham, 2017:3-17.

[10] Harer, Jacob A., et al. "Automated software vulnerability detection with machine learning." arXiv preprint arXiv:1803.04497(2018).

[11] Du, Xuedan , et al. "Overview of deep learning." Chinese Association of Automation IEEE, 2017:27-51.

[12] Deng, Jia, et al. "ImageNet: A large-scale hierarchical image database." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on IEEE, 2009:248-255.

[13] Russakovsky, Olga, et al. "ImageNet Large Scale Visual Recognition Challenge." International Journal of Computer Vision 115.3(2015):211-252.

[14] Xiong, W, et al. "Achieving Human Parity in Conversational Speech Recognition." IEEE/ACM Transactions on Audio Speech & Language Processing, 2016: 99.

[15] Xiong W, Droppo J , Huang X , et al. Toward Human Parity in Conversational Speech Recognition[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(12):2410-2423.

[16] Krizhevsky, Alex, I. Sutskever, and G. E. Hinton. "ImageNet classification with deep convolutional neural networks." International Conference on Neural Information Processing Systems Curran Associates Inc. 2012:1097-1105.

[17] Liu, Xuan , et al. "Deep convolutional neural networks-based age and gender classification with facial images." First International Conference on Electronics Instrumentation & Information Systems IEEE, 2018: 1-4.

[18] Silver, D, et al. "Mastering the game of Go with deep neural networks and tree search." Nature, 2016, 529(7587):484-489.

[19] Iyyer, Mohit, et al. "Deep Unordered Composition Rivals Syntactic Methods for Text Classification." Meeting of the Association for Computational Linguistics and the, International Joint Conference on Natural Language Processing, 2015, 1681-1691.

[20] Jo, Hwiyeol, et al. "Large-Scale Text Classification with Deep Neural Networks." Korea Computer Congress, 2016.

[21] Aziguli, Wulamu, et al. "A Robust Text Classifier Based on Denoising Deep Neural Network in the Analysis of Big Data." Scientific Programming, 2017:1-10.

[22] Wu, Fang , et al. "Vulnerability detection with deep learning." IEEE International Conference on Computer & Communications IEEE, 2018: 1298-1302.

[23] Li, Zhen , et al. "VulDeePecker: A Deep Learning-Based System for Vulnerability Detection." arXiv preprint arXiv:1801.01681 (2018).

[24] Zhang, Hao, K. Lv, and C. Hu. "An Automatic Vulnerabilities Classification Method Based on Their Relevance." International Conference on Network and System Security Springer, Cham, 2017:475-485.

[25] Berrar, Daniel, and W. Dubitzky. "Information Gain." Springer New York, 2013.

[26] National Vulnerability Database. http://nvd.nist.gov/

[27] Stop word list. https://pypi.org/project/stop-words/

[28] Turney P D , Pantel P. "From Frequency to Meaning: Vector Space Models of Semantics." Journal of Artificial Intelligence Research, 2010, 37(1):141-188.

[29] Sun A, Lim E P, Liu Y. "On strategies for imbalanced text classification using SVM: A comparative study." Decision Support Systems, 2009, 48(1):191-201.

[30] Shi, Kansheng, et al. "An improved KNN text classification algorithm based on density." IEEE, International Conference on Cognitive Informatics & Cognitive Computing IEEE Computer Society, 2017:92-95.