A data enhancement method for security classification of intelligent terminal

Zhining Lv [1], Jianfang Song[2], Wei Deng [1], Zhongwei Qiao [1], Yi Chen[2], Yu Du [1], Hong Wen[2]

[1.] Shenzhen Power Supply Bureau Co, Ltd., Shenzhen 510820, China

[2.] School of Aeronautics and Astronautics, University of Elec. Science and Tech. of China, Chengdu 611731, China

sunlike@uestc.edu.cn

Keywords: Data Enhancement, Intelligent terminal, Grade Assessment, Security Classification

Abstract: With the perfect combination of intelligent terminal and internet. Smart terminals increasingly involve sensitive information such as business secrets and personal privacy. So, it is very necessary to have a model for assessing the security performance of intelligent terminals. Current methods of terminal classification, especially artificial intelligence (AI) technology, which have the problems of long test time and insufficient precision because of insufficient data. In view of this problem, this project proposed a data enhancement method for security classification of intelligent terminals. The article first introduces the method of acquiring enhanced data sets, and then uses the SVM algorithm to implement the hierarchical classification of smart terminals based on the acquired enhanced data sets. The randomness of samples can be enhanced, so the robustness of samples can be enhanced. It is of great significance to the research of terminal security classification.

1. Introduction

With the popularization of the network and the development of 4G / 5G wireless network, the application of terminals has penetrated into our daily life. However, compared with PC, the terminal is more vulnerable to attack due to its limited size, energy and computing power, and the terminal is widely distributed in personal and various application scenarios and easy to access, which makes it more and more vulnerable to attack ([1], [7]). As people's dependence on intelligent terminal, the terminal security problem is also increasingly prominent. In particular, the personal privacy [2] and commercial confidential documents stored in the intelligent terminal. In recent ten years, a variety of attacks [3] on mobile intelligent terminals emerge one after another. The attacks on terminals become an entry point to attack the network, and the hidden dangers of terminals also become an important issue of network security. In this context, it is necessary to evaluate the security of intelligent terminals.

In mobile intelligent terminal security assessment, Mobile intelligent terminal security evaluation has become one of the most effective means to ensure the safe use of intelligent terminal, which can achieve the objective and accurate classification of the security level of mobile intelligent terminal [4]. At present, regarding the quantization division of terminal safety level, the methods of terminal single safety test have achieved certain results ([5], [8]), and then, by synthesizing some quantitative data of terminal security performance, advanced classification method is adopted for terminal security classification, especially based on artificial intelligence (AI) technology, through the learning algorithm, the security performance of the terminal can be objectively classified. However, the classification method based on artificial intelligence (AI) technology requires a large

  DOI: 10.25236/icicta.2019.016

amount of data to train the model, and it takes a long time to test the data [6]. If the data is insufficient, the classification accuracy will not be accurate enough. In view of this problem, this project proposes a data enhancement method for smart terminal security level classification.

2. Terminal Security Classification Based on Data Enhancement Method

The enhanced data set is obtained through the data enhancement method, and then the enhanced data set is used as training data to classify the security level of the intelligent terminal based on the SVM algorithm.

2. 1 Get the enhanced dataset

The purpose of this scheme is to overcome the shortcomings of existing technology, to provide a data enhancement method for intelligent terminal security classification, to construct new pseudo-evaluation samples by using the correlation of terminal security evaluation samples, and to introduce the concept of random weight to increase the randomness of sample construction, so as to enhance the robustness of sample set.

Get enhanced datasets through data augmentation methods, including the following steps:

S1: Test the kth intelligent terminal for S times and get the test result $\boldsymbol{M_k^1}, \boldsymbol{M_k^2}, \cdots, \boldsymbol{M_k^S}$, each test result consists of n individual test scores, that is represented by $\boldsymbol{M_k^i} = [m_1, m_2, \cdots, m_n]^T$, where m_j is the score of the jth single test case, the higher the score, the better the security performance [10];

S2: Multiply the test result by the weight function $H(n)$ of each single instance to get the total score Y of each terminal. Where the weight function $H(n)$ is a uniform probability density function, expressed as $H = [h_1, h_2, \cdots, h_S]^T$, $h_i = \frac{1}{N}$, i.e.$Y = (M_i^k)^T.H$; At the same time, the terminal security level y is divided into W levels, and $W-1$ thresholds are set to be a positive number $\eta_1, \eta_2, \cdots, \eta_{W-1}$, When $0 < Y \le \eta_1$ is satisfied, the terminal security level is defined as level 1, and when $\eta_1 < Y \le \eta_2$ is satisfied, the security level is defined as level 2, and by analogy, when $\eta_{W-2} < Y \le \eta_{W-1}$ is satisfied, the security level of the terminal is defined as $K-1$; when $Y > \eta_{W-1}$ is satisfied, the security level is defined as W. The higher the security level, the more secure the terminal;

S3: By testing the calculation of the total score M_i and the security level y of a terminal, the S times test data set $\boldsymbol{D}_K$ of the kth terminal is:

$$\boldsymbol{D}_k : \boldsymbol{D}_k = \{\boldsymbol{X}_k, \boldsymbol{Y}_k\} \tag{1}$$

Where $\boldsymbol{X}_k = \left[\boldsymbol{M}_k^1, \boldsymbol{M}_k^2, \cdots, \boldsymbol{M}_k^S\right]$, and $T = \{(M_1, y_1), (M_2, y_2), \cdots, (M_N, y_N)\}$, $y_i \in \{1, 2, 3, 4\}$, $i = 1, 2, \cdots, N$;

S4: The corresponding labels (node number), that is, the output sample set is:

$$\boldsymbol{Y}_k = \underbrace{\left[\boldsymbol{y}_k, \boldsymbol{y}_k, \cdots, \boldsymbol{y}_k\right]}_{S} \tag{2}$$

Where $y_k \in \{1,2,\cdots,W\}$;

S5: Then, construct a new input channel information sample according to the following formula:

$$\boldsymbol{M}_k^n = \frac{1}{\alpha_0 + 1}\sum_{i=n}^{n+\infty} \boldsymbol{M}_k^i, 1 \le \alpha_0 < S, n + \alpha_0 \le S \tag{3}$$

Where α_0 is a positive integer representing the number of samples constructed for each parameter evaluation sample;

S6: The input sample set after average data enhancement is:

$$\boldsymbol{X}_k^{\dagger} = \left[\boldsymbol{M}_k^1, \boldsymbol{M}_k^2, \cdots, \boldsymbol{M}_k^S, \boldsymbol{M}_k^1, \cdots, \boldsymbol{M}_k^{N_k}\right] \tag{4}$$

Among them, N_k represents the number of channel information vectors after average data enhancement;

S7: The label matrix after average sample construction, that is, the output sample set is:

$$\boldsymbol{Y}_k^{\dagger} = \underbrace{\left[\boldsymbol{y}_k, \boldsymbol{y}_k, \cdots, \boldsymbol{y}_k\right]}_{S+N_k} \tag{5}$$

S8: Get enhanced dataset:

$$\boldsymbol{\mathcal{D}}_k^{\dagger}: \boldsymbol{\mathcal{D}}_k^{\dagger} = \{\boldsymbol{X}_k^{\dagger}, \boldsymbol{Y}_k^{\dagger}\} \tag{6}$$

The process of obtaining an enhanced data set is illustrated below with an actual data set.

This experiment is based on the established security test development platform, taking the evaluation of the mobile smart terminal model as Xiaomi 4C, the system is Android 6.1.1 as an example, testing from four aspects of system security, storage security, privacy security and application security. The test safety from low to high scores range from 0-100.

For example: the first test result of a terminal is $M_k^i = [m_1, m_2, \cdots m_n] = [95,50,80,95]$, Among them, $i = 5$, $n = 4$, and get 5 times test results of one terminal in turn:

$$\boldsymbol{X_k} = [\boldsymbol{M}_K^1, \boldsymbol{M}_K^2, \cdots, \boldsymbol{M}_K^S] = \begin{bmatrix} 95 & 95 & 100 & 95 & 100 \\ 50 & 60 & 50 & 60 & 50 \\ 80 & 85 & 85 & 85 & 85 \\ 95 & 100 & 95 & 95 & 95 \end{bmatrix} \tag{7}$$

A new input channel information sample constructed by formula (3), where $\alpha_0 = 1$, and then, the input sample set after average enhancement is:

$$\boldsymbol{X}_k^{\dagger} = \left[\boldsymbol{M}_K^1, \boldsymbol{M}_K^2, \cdots, \boldsymbol{M}_K^S, \boldsymbol{M}_K^1, \cdots, \boldsymbol{M}_K^{N_k}\right] = \begin{bmatrix} 95 & 95 & 100 & 95 & 100 & 97.5 \\ 50 & 60 & 50 & 60 & 50 & 55 \\ 80 & 85 & 85 & 85 & 85 & 85 \\ 95 & 100 & 95 & 95 & 95 & 95 \end{bmatrix} \tag{8}$$

Get test item weight vector $H(n) = [0.373 \ \ 0.163 \ \ 0.278 \ \ 0.185]$ based on AHP algorithm, and then, the output enhanced sample set is obtained.

$$\boldsymbol{Y}_k^{\dagger} = \underbrace{[\boldsymbol{I_k}, \ \boldsymbol{I_k}, \ \cdots, \ \boldsymbol{I_k}\ ]}_{S+N_k} = [79.95 \ \ 84.47 \ \ 82.63 \ \ 83.54 \ \ 82.63 \ \ 83.09]^T \tag{9}$$

Where $N_k = 1$.

S9: Using the new data set obtained by formula (9) as the training set, the AI-based terminal security level classifier is trained.

The beneficial effects of this solution are:

(1) A new pseudo-evaluation sample is constructed by using the correlation of the terminal security evaluation samples, and the concept of random weight is introduced to increase the randomness of the sample structure to enhance the robustness of the sample set.

(2) The data collection enhancement method of this solution can be applied to the enhancement of various AI-based terminal security level classifier data.

(3) This security level classification data collection enhancement method can be applied to a variety of different intelligent terminal devices, has strong portability, and is widely used.

2.2 Classification of Terminal Security Level Based on SVM Algorithm

Taking the support vector machine (SVM) terminal security level classification method as an example ([9], [11]), the enhanced data described in this solution S1-S9 for the security level classification of the SVM method includes the following steps.

According to the level W of the security level, a $W-1$ layers support vector machine model is used to calculate the security level, including the following sub-steps [12]:

(1) Initialize, make the initial variable $m=1$;

(2) Divide the training set into two classes, where one class is $y=m$ and another class level is $y=m+1\sim W$, that is, training set $\boldsymbol{\mathcal{D}}_k^{\dagger}$: $\boldsymbol{\mathcal{D}}_k^{\dagger}=\{\boldsymbol{X}_k^{\dagger},\boldsymbol{Y}_k^{\dagger}\}$ is obtained, where $\boldsymbol{X}_k^{\dagger}$ and $\boldsymbol{Y}_k^{\dagger}$ is shown by equations (4) and (5), and $\boldsymbol{y}_k$ in formula (5) is $y_k=\begin{cases}1, & y_k=1;\\ -1, & y_k\in\{2,\cdots,W\};\end{cases}\ k=1,2,\cdots,S+N_k$;

(3) Construct and solve the constraint optimization problem, the formula is as follows:

$$\min_{\alpha}\ \frac{1}{2}\sum_{i=1}^{S+N_k}\sum_{j=i}^{S+N_k}\alpha_i\alpha_j y_i y_j(x_i\bullet x_j)-\sum_{i=1}^{S+N_k}\alpha_i \tag{10}$$

$$s.t.\ \sum_{i=1}^{S+N_k}\alpha_i y_i=0,\ \ \alpha_i\geq 0,\ \ i=1,2,\cdots,S+N_k \tag{11}$$

Find the optimal solution $\alpha^{(m)}=(\alpha_1^{(m)},\alpha_2^{(m)},\cdots,\alpha_{S+N_k}^{(m)})^T$, where $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_{S+N_k})^T$ is the Lagrange multiplier vector, $x_i\in\chi=R^n$, $y_i\in\gamma=\{+1,-1\}$, $i=1,2,3,\cdots,S+N_k$;

(4) Calculate:

$$w^{(m)}=\sum_{i=1}^{S+N_k}\alpha_i^{(m)}y_i x_i \tag{12}$$

In the formula, w represents the normal vector value of the classification hyperplane in the high-dimensional space; at the same time, select a positive component of $\alpha_j^{(m)}>0$, and calculate

$$b^{(m)}=y_j-\sum_{i=1}^{S+N_k}\alpha_i^{(m)}y_i(x_i\bullet x_j) \tag{13}$$

Where b is the intercept value of the classification hyperplane in the high-dimensional space;

(5) Get the hyperplane by calculation $w^{(m)}\bullet x+b^{(m)}=0$, and then, by the classifying decision function

$$f^{(m)}(x) = sign(w^{(m)} \bullet x + b^{(m)}) \tag{14}$$

The terminal with security level m is divided: When $f^{(1)}(M_i) = 1$, the terminal security level is m, and when $f^{(1)}(M_i) = -1$, the terminal security level is $m+1 \sim W$;

(6) Determine if the value of m is equal to $w-1$: if yes, complete all security level classifications; if not, perform +1 operation on m, and return to step (2).

The test items include text message function, call function, third-party software, kernel vulnerability, audit function, storage and deletion of file warnings. The security level is 4 levels, where 3 thresholds need to be set, and a 3-layer support vector machine model is used.

The flowchart of the method implemented in this solution is as follows.

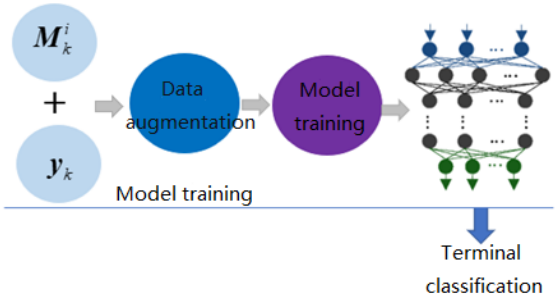

Figure1. Method flow for technical solution implementation

3. Conclusion

Aiming at the problem of long test data and insufficient accuracy caused by insufficient data based on the AI algorithm security level classification method, this paper proposes a data enhancement method for security classification of intelligent terminal. This method overcomes the shortcomings of the existing technology, enhances the robustness of the sample set, and realizes different security requirements of users on smart terminals.

Acknowledgements

This work is supported by National major R&D program (2018YFB0904900, 2018YFB0904905).

References

[1] Qing Yang, Yixin Jiang, Aidong Xu, Hong Wen, Feng Wang, LiuFei Chen, Kai Ouyang, Xinping Zhu, “A Model Divides the Mobile Security Level Based on SVM,” in Proceedings of IEEE CNS 2017, LasVegas, USA, 7-9 Oct., 2017.

[2] Q. Jia, L. Guo, Z. Jin, et al. Privacy-preserving data classification and similarity evaluation for distributed systems, 2016 IEEE 36th International Conference on Distributed Computing Systems, 2016, 690-699.

[3] Idrees F , Rajarajan M , Conti M , et al. PIndroid: A novel Android malware detection system using ensemble learning methods[J]. Computers & Security, 2017, 36-46.

[4] G Wang, Q Song, X Zhu. An improved data characterization method and its application in classification algorithm recommendation, Applied Intelligence, 2015, 43(04):892-912.

[5] Xu Jiang,Lili Liu,Simeng Yu,Guohui Lyu,Xiaohan Zhan. Research on a monitoring terminal for a fibre grating sensing device based on Android[J]. Pacific Science Review. 2014 (1).

[6] Jie Tang, Hong Wen, Kai Zeng, Run-fa Liao, Fei Pan, Lin Hu, Light-weight physical layer enhanced security schemes for 5G Wireless Networks, IEEE Network. Volume: 33, Issue: 5, pp. 126 – 133, Sep. 2019.

[7] Dagon C, Martin T, Starner T. Mobile Phones as Computing Devices: The Viruses Are Coming[J]. IEEE Pervasive Computing,2004(03):11-15.

[8] S. Saychum, A. Thangthai, P. Janjoi, et al. A bi-lingual Thai-English TTS system on Android mobile devices[C]. IEEE International Conference on Electrical Engineering/electronics, Computer, Telecommunications and Information Technology. 2012:1-4.

[9] Y. Hai, W. He, L. Fan. An incremental learning algorithm for SVM based on voting principle[J].

[10] Singh S K, Mishra B, Gera P. A Privacy Enhanced Security Framework for Android Users[C]. IEEE International Conference on It Convergence and Security, 2015:1-6.

[11] J. Lin, M. Song, J. Hu. An SMO approach to fast SVM for classification of large scale data[C]. IEEE International Conference on It Convergence and Security. 2014:1-4.

[12] Yang Y M.Decomposition and recognition of playing volleyball action based on SVM algorithm[J]. Journal of Interdisciplinary Mathematics, 2018, 21 (5) :1181-1186.

[13] International Journal of Information Processing & Management, 2010, 2(2):420-423.