Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2020.010401
ISSN 2790-0983 Vol. 1, Issue 4: 1-8

# *Sensor Cloud Intrusion Detection Based on Discrete Optimization Algorithm and Machine Learning*

**Xiaoqi Yin**[*]

*Heilongjiang University of Industry and Business, Harbin 150025, China*

*42159919@qq.com*

[*]*corresponding author*

***Keywords:*** Discrete Optimization Algorithm, Machine Learning, Sensor Cloud, Intrusion Detection

***Abstract:*** In the development process of the Internet, computer technology and network communication have been rapidly applied. Network security has become a research focus. Based on the discrete cloud intrusion detection method, through sensing a large number of data signals in the cloud environment, some of the interference information is screened, filtered and classified. In this paper, the discrete optimization algorithm and machine learning can better reflect the intrusion information in time. This paper mainly uses the methods of experiment and comparison to experiment the three indicators of SVM and its improved algorithm in intrusion detection. The experimental data show that the accuracy of the improved LE-SVM algorithm can reach more than 95%, and its time consumption is relatively small.

## 1. Introduction

With the rapid development of the Internet and the rapid improvement of computer technology, people's lives have been inseparable from the Internet. However, with the growing demand for online communication services and meeting the use requirements of a large number of users, cloud services have been valued. The main function of the sensor network is to transmit data information through collection, storage and processing, and detect the content contained in the transmission process, so as to achieve effective filtering of traffic and other relevant parameters. The first thing to do in cloud intrusion based on discrete optimization algorithm is to simulate the signal flow environment. Discrete algorithm can play a great role in cloud intrusion detection.

There are many researches based on discrete optimization algorithm and machine learning in sensor cloud intrusion detection. For example, some people proposed a sensor cloud intrusion detection algorithm based on parallel discrete optimization feature extraction for large-scale high-dimensional data and variable intrusion behavior of the sensor cloud [1-2]. Some people propose a sensor cloud intrusion detection method based on machine learning to solve the problem that sensor cloud networks are vulnerable to intrusion attacks [3-4]. Others say that intrusion detection system is a kind of monitoring of network transmission data, system logs, and system user activities [5-6]. Therefore, the research on intrusion system needs to focus on the improvement of technology. In this paper, discrete optimization algorithm and machine learning can play a huge role in sensor cloud intrusion detection.

In this paper, discrete particle swarm optimization algorithm is first studied, and its iterative process is described in detail. Secondly, the form of machine learning is discussed, and the importance of data processing in sensor cloud intrusion detection is obtained. Then the wireless sensor intrusion detection technology is discussed in detail. Finally, through several improved algorithms, the experimental simulation draws relevant data and conclusions.

## 2. Sensor Cloud Intrusion Detection Based on Discrete Optimization Algorithm and Machine Learning

### 2.1. Discrete Particle Swarm Optimization Algorithm

By using the discrete particle swarm optimization algorithm, we can use the parallel updating strategy of particles to solve the complex network intrusion detection problem.

First, the network data is loaded into the distributed computing platform in the form of adjacency matrix, and broadcast variables are used to enable all servers in the cluster to obtain this matrix. In this algorithm, the module density is used as the objective function to calculate the new fitness of each particle after each location update. The search path of particles is guided by the global optimal solution and local optimal solution in the particle swarm optimization algorithm, and finally the particles with the largest module density are obtained. Decode the position of particles corresponding to this result to obtain community division [7-8].

The iterative process of parallel discrete particle swarm optimization algorithm is as follows.

Firstly, the speed and location of particles are updated in parallel by taking advantage of the parallelization of elastic distributed data sets, and the global optimal location is shared among servers as a broadcast variable. Therefore, a single particle has all the necessary information in the iteration process, so that all particles can be updated at the same time. In the same case, the adjacency matrix is shared among multiple servers in the form of broadcast variables, so that the process of computing fitness for each particle can also be operated in parallel. The second stage is to obtain the global optimal location. Since each particle is scattered in multiple locations of the whole cluster, the Reduce operation of distributed dataset is used to summarize the optimal location. A suitable intrusion partition can be finally obtained through the iteration of the parallel particle swarm optimization algorithm [9-10].

### 2.2. Machine Learning Form

In this paper, we use supervised learning to preprocess the training set features and labels. Then, the machine learning algorithm is used to train, and the trained model is used to judge whether an access record has intrusion behavior. Machine learning is mainly artificial, using computers as information processing tools to solve problems through imitation, association and other ways. Machine learning is to code, calculate, express and propagate specific programs or processes by imitating the structure and function of computers and using existing tools. It has been widely used in many fields. In sensor networks, data acquisition is done manually. Machine learning can integrate a large amount of redundant information for processing. In sensor networks, data transmission and processing are very important and difficult. Therefore, it is necessary to use machine learning to complete intrusion detection of sensors. In sensor networks, a large amount of information will be lost or destroyed due to complexity and difficulty in storage. Therefore, when collecting data from sensors, it is necessary to use machine learning methods to train samples [11-12].

The sensor cloud intrusion detection based on machine learning is based on a trained data stream, which interacts with sub modules (database management and query). Each module has its own specific function implementation rules. Each part can be combined into a group of output data and sent to the microcontroller as input data. The device enters the global optimal search range after executing the corresponding command. The return program stores these results in the current machine learning platform and calls the local optimization algorithm to determine whether there is an intrusion [13-14].

### 2.3. Wireless Sensor Intrusion Detection Technology

Intrusion detection system is a collection of related hardware and software for intrusion detection, which is called IDS for short. Nowadays, intrusion detection systems have been widely used in the field of network security to solve the security problems of computer networks and host systems. The integration of network monitoring and network management functions has become a development trend. Intrusion detection technology can detect data packets in the network, and immediately process and manage the abnormal node after it is found. In the future, the intrusion detection system will integrate some relevant network management software to build a tool that can perform intrusion detection, monitor the network environment and manage the entire network [15-16].

Traditional intrusion detection is mainly based on artificial neural network and fuzzy pattern recognition. These two methods are relatively mature and widely used at present. But for sensor networks, data preprocessing is the first step. Because of its complex working environment, high requirements for data information and a certain degree of concealment, it will affect the actual detection effect. Machine learning can effectively implement intrusion detection and location in virtual space. Secondly, a large number of training samples are required, and at the same time, good implicit ability and learning characteristics are required to build the model and determine the attack target. Intrusion detection technology of cloud services is an important guarantee to prevent and prevent various risks and hazards. Among them, the sensor system has the largest amount of data, that is, the critical coefficient is the smallest. Therefore, how to improve intrusion detection technology to attack various types of information is particularly critical. At present, it is often used to analyze and process data based on discrete optimization algorithm research methods [17-18].

## 3. System Structure Design

### 3.1. Overall System Design

The overall structure of the intrusion detection system is shown in Figure 1:

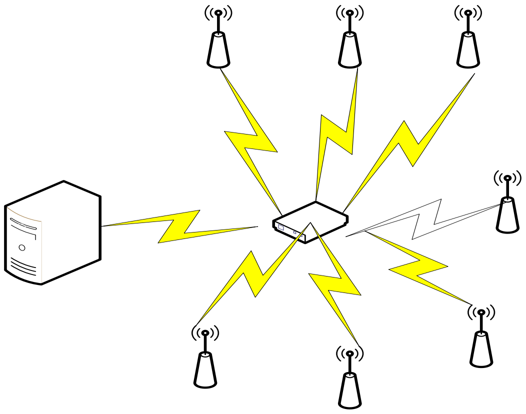

*Figure 1. Overall structure of the hardware system*

The first part is the outermost part, called monitoring node, which is distributed near the boundary of the entire range area. As the peripheral monitoring sensor node of the boundary intrusion detection system, it is responsible for monitoring whether foreign objects invade the range area. The second part is the central part of the sensor network, called the sink node. As the center of the intrusion detection system, the sink node is responsible for receiving the information sent by the monitoring node and sending sensor data to the monitoring host through the serial port. The third part is the monitoring host, which is responsible for receiving the data sent by the sensor and data fusion through computer software.

## 3.2. Experimental Environment

The experimental environment and data set used in this experiment are the same as those of the same experiment, namely: Windows 13 operating system, Matlab 8.12.0, and the hardware environment processor is Core 4, 2.50GHz, 5G RAM and KDD Cup data set.

## 3.3. Experimental Data and Scheme

In this experiment, 2000 connection records were randomly extracted as training samples, and another 2000 connection records were randomly extracted as test samples, each containing 1500 normal data and DOS / Probing containing 80 connection records, respectively, and R2L / U2R containing 60 connection records, respectively. The continuous numerical values were normalized. The modified MDS- -GA- -SVM algorithm was then applied to intrusion detection.

Use MDS to reduce the dimension of the preprocessed dataset. SVM is used to classify data sets, and genetic algorithm is used to optimize the two parameters in SVM. The parameters of the improved MDS-GA-SVM are set according to many experiences. The values of continuous type are normalized:

$$\mathrm{w}_{\mathrm{n}} = \frac{w - q_{\min}}{q_{\max} - q_{\min}} \tag{1}$$

Where represents the normalized value. The eigen dimension of the data set can be estimated to be 6 by the maximum likelihood estimation method, and then the LE algorithm is used to reduce the dimension of the data set. The SVM intrusion detection classifier is constructed by using LIBSVM function library to classify the reduced dimension data. Analyze which algorithm is the best by observing the accuracy, false alarm rate and detection time. The formula of false alarm rate is as follows:

$$E = \frac{E_{number}}{N_{number}} \times 100\% \tag{2}$$

Among them, E represents the false alarm rate. The detection time represents the time taken for toc to output the experimental process by adding tic to the experiment.

## 4. Simulation Result Analysis

### 4.1. Simulation Experiment Results of U2R, DOS, Probing and R2L

After many experiments, it is proved that LE is a manifold learning algorithm with the fastest convergence speed, and LE has a complete spectrum theory, so this algorithm selects LE algorithm. Finally, the simulation experiment results of R2L accuracy, false alarm rate and detection time obtained by SVM algorithm, C-SVM algorithm and LE-SVM algorithm are shown in Table 1:

*Table 1. Results of the R2L simulation experiments*

| | SVM | C-SVM | LE-SVM |
|---|---|---|---|
| Accuracy rate | 88.9 | 95.2 | 95.8 |
| Error rate | 2.8 | 1.78 | 1.71 |
| Detection time | 1.76 | 19 | 5.78 |

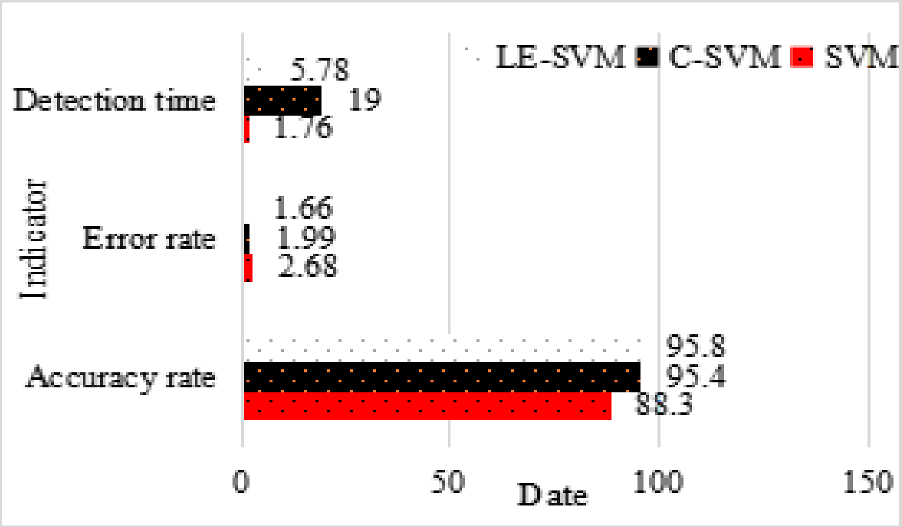

*Figure 2. Results of the U2R simulation experiments*

As shown in Figure 2, in terms of accuracy, the accuracy of classification using only SVM is low, while the accuracy of C-SVM and LE-SVM is high. C-SVM and LE-SVM are both excellent in accuracy.

In terms of false alarm rate, the false alarm rate of the three algorithms is about 2%. In terms of detection time, it can be seen from Table 2 that the detection time of CV-SVM is the longest and that of SVM is the shortest. The detection time of LE SVM is slightly more than that of SVM, but the time spent is one-third of that of C-SVM.

*Table 2. Results of the DOS simulation experiments*

| | SVM | C-SVM | LE-SVM |
|---|---|---|---|
| Accuracy rate | 89.7 | 96.1 | 96.7 |
| Error rate | 2.3 | 1.9 | 1.8 |
| Detection time | 1.76 | 19 | 5.78 |

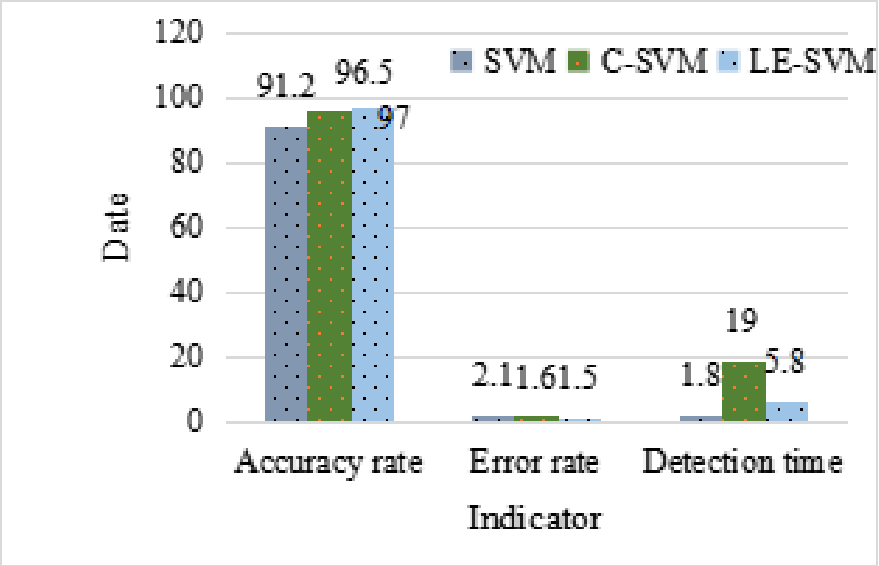

*Figure 3. Results of the probing simulation experiments*

As shown in Figure 3, the accuracy rate of the LE-SVM algorithm is slightly higher than that of the C-SVM algorithm. The accuracy rates of both algorithms are very high. In terms of detection time, the SVM algorithm and the LE-SVM algorithm are significantly better than the C-SVM algorithm. To sum up, the LE-SVM algorithm proposed in this paper has three indicators: comprehensive accuracy, detection time, and false alarm rate. The LE-SVM algorithm is superior to the other two algorithms, realizing the characteristics of short training time, high accuracy, and low false alarm rate.

## 5. Conclusion

In traditional sensor cloud intrusion detection methods, a large amount of data needs to be counted. In addition, when invading, the signals acquired by sensors should be classified and processed first. However, this will lead to a longer running time of the algorithm and an increase in the amount of computation. Now we use discrete optimization technology to achieve signal analysis and extraction tasks. Through the research in this paper, the discrete algorithm can more quickly and accurately fit the feature pattern recognition, and through the machine learning model, it can occupy an advantage in data processing. In addition, the three optimized clustering algorithms mentioned in this paper can also process data information well.

## Funding

This article is not supported by any foundation.

## Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

## Conflict of Interest

The author states that this article has no conflict of interest.

## References

[1] *Hüseyin Hakli, Harun Uguz, Zeynep Ortacay: Comparing the Performances of six Nature-Inspired Algorithms on a Real-World Discrete Optimization Problem. Soft Comput. 26(21): 11645-11667 (2020).*

[2] *Seung-Yeal Ha, Shi Jin, Doheon Kim: Convergence and Error Estimates for Time-Discrete Consensus-Based Optimization Algorithms. Numerische Mathematik 147(2): 255-282 (2020).*

[3] *Hadi Jahangir, Mohammad Mohammadi, Seyed Hamid Reza Pasandideh, Neda Zendehdel Nobari: Comparing Performance of Genetic and Discrete Invasive Weed Optimization Algorithms for Solving the inventory Routing Problem with an Incremental Delivery. J. Intell. Manuf. 30(6): 2327-2353 (2019). https://doi.org/10.1007/s10845-018-1393-z*

[4] *Sedigheh Mahdavi, Shahryar Rahnamayan, Abbas Mahdavi: Majority Voting for Discrete Population-Based Optimization Algorithms. Soft Comput. 23(1): 1-18 (2019). https://doi.org/10.1007/s00500-018-3530-1*

[5] *Goutam Sen, Mohan Krishnamoorthy: Discrete particle Swarm Optimization Algorithms for Two Variants of the Static Data Segment Location Problem. Appl. Intell. 48(3): 771-790 (2017). https://doi.org/10.1007/s10489-017-0995-z*

[6] *David A. Brown, Siva Nadarajah: Inexactly constrained discrete adjoint approach for steepest descent-based optimization algorithms. Numer. Algorithms 78(3): 983-1000 (2017). https://doi.org/10.1007/s11075-017-0409-7*

[7] *Simon Birnbach, Richard Baker, Simon Eberz, Ivan Martinovic: #PrettyFlyForAWiFi: Real-world Detection of Privacy Invasion Attacks by Drones. ACM Trans. Priv. Secur. 24(4): 31:1-31:34 (2020). https://doi.org/10.1145/3473672*

[8] *Benjamin Doerr, Frank Neumann: A Survey on Recent Progress in the Theory of Evolutionary Algorithms for Discrete Optimization. ACM Trans. Evol. Learn. Optim. 1(4): 16:1-16:43 (2020). https://doi.org/10.1145/3472304*

[9] *Habeeb Bello-Salau, Adeiza J. Onumanyi, Adnan M. Abu-Mahfouz, Achonu Adejo, Muhammed Bashir Mu'azu: New Discrete Cuckoo Search Optimization Algorithms for Effective Route Discovery in IoT-Based Vehicular Ad-Hoc Networks. IEEE Access 8: 145469-145488 (2020). https://doi.org/10.1109/ACCESS.2020.3014736*

[10] *Peter L. Salemi, Eunhye Song, Barry L. Nelson, Jeremy Staum: Gaussian Markov Random Fields for Discrete Optimization via Simulation: Framework and Algorithms. Oper. Res. 67(1): 250-266 (2018). https://doi.org/10.1287/opre.2018.1778*

[11] *Hossein Safi, Mohammad Ali Montazeri, Javane Rostampoor, Saeedeh Parsaeefard: Spectrum Sensing and Resource Allocation for 5G Heterogeneous Cloud Radio Access Networks. IET Commun. 16(4): 348-358 (2020). https://doi.org/10.1049/cmu2.12356*

[12] *Sorour Mohajerani, Parvaneh Saeedi: Pragmatic Augmentation Algorithms for Deep Learning-Based Cloud and Cloud Shadow Detection in Remote Sensing Imagery. IEEE Geosci.*

*Remote. Sens. Lett. 19: 1-5 (2020).*

*[13] Ramon Padullés, Estel Cardellach, F. Joseph Turk, Chi O. Ao, Manuel de la Torre Juarez, Jie Gong, Dong L. Wu: Sensing Horizontally Oriented Frozen Particles With Polarimetric Radio Occultations Aboard PAZ: Validation Using GMI Coincident Observations and Cloudsat a Priori Information. IEEE Trans. Geosci. Remote. Sens. 60: 1-13 (2020).*

*[14] Sanaz Kianoush, Stefano Savazzi, Manuel Beschi, Stephan Sigg, Vittorio Rampa: A Multisensory Edge-Cloud Platform for Opportunistic Radio Sensing in Cobot Environments. IEEE Internet Things J. 8(2): 1154-1168 (2020). https://doi.org/10.1109/JIOT.2020.3011809*

*[15] Naga Raju Hari Manikyam, Munisamy Shyamala Devi: A Framework for Leveraging Image Security in Cloud with Simultaneous Compression and Encryption Using Compressive Sensing. Rev. d'Intelligence Artif. 35(1): 85-91 (2020). https://doi.org/10.18280/ria.350110*

*[16] Alex Paludo, Willyan Ronaldo Becker, Jonathan Richetti, Laíza Cavalcante De Albuquerque Silva, Jerry Adriani Johann: Mapping Summer Soybean and Corn with Remote Sensing on Google Earth Engine Cloud Computing in Parana State - Brazil. Int. J. Digit. Earth 13(12): 1624-1636 (2020). https://doi.org/10.1080/17538947.2020.1772893*

*[17] B. Maheswara Rao, S. Baskar: Enhanced Modulation Scheme for Cognitive Radio over Rayleigh Fading Channels Using Power Allocation and Spectrum Sensing Models. Int. J. Cloud Comput. 9(2/3): 285-294 (2020). https://doi.org/10.1504/IJCC.2020.10031546*

*[18] Meisam Amani, Arsalan Ghorbanian, Seyed Ali Ahmadi, Mohammad Kakooei, Armin Moghimi, S. Mohammad MirMazloumi, Sayyed Hamed Alizadeh Moghaddam, Sahel Mahdavi, Masoud Ghahremanloo, Saeid Parsian, Qiusheng Wu, Brian Brisco: Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 13: 5326-5350 (2020). https://doi.org/10.1109/JSTARS.2020.3021052*