Support Vector Machine with Improved Particle Swarm Optimization Model for Intrusion Detection

S P Yazhini[1], Dr. R Devipriya[2]

[1]PG Student, Department of Information Technology, Kongu Engineering College, Tamil Nadu, India.
E-mail: yazhisivan@gmail.com

[2]Assistant Professor,Department of Information Technolog, Kongu Engineering College,Tamil Nadu, India.
E-mail: scrpriya@gmail.com

ABSTRACT

Intrusion Detection System (IDS) is a computer-based data system which purports to observe attacks against computer systems and networks or, against any information system. Its job is to supervise the utilization of such system to detect any insecure states. IDS detect attempts and active misuse of the scheme either by lawful users of the information systems or by outside parties to abuse privileges or exploit security vulnerabilities. It gets information about target system to perform diagnosis on security status. Data mining techniques used for intrusion detection are classification, clustering, frequent pattern mining and mining data streams. The classification method called Support Vector Machine (SVM) has been used to provide potential results for the intrusion detection problem. Nevertheless, the practicability of SVM is affected due to the trouble of selecting appropriate SVM parameters and feature selection. The optimization algorithm, Particle Swarm Optimization (PSO) is applied to pick out the optimized parameters for the SVM. For initializing the population of PSO, Optimal Latin Hypercube Design (OLHD) is applied. The OLHD follows the space-filling property of attributes in the design space. The proposed OLCPSO-SVM model is employed to an intrusion detection problem in the KDD Cup 99 data set. The experimental results show that the OLCPSO-SVM method can reach a higher detection rate than regular SVM algorithms.

Key Words: Intrusion detection system, Particle swarm optimization, Support vector machine, Optimal latin hypercube design

1. INTRODUCTION

In an information system, intrusions are the cognitive operations that violate the security policies of the system and intrusion detection is the operation used to identify intrusions. An intrusion detection system is a type of security software designed to automatically alert administrators when someone or something is trying to compromise an information system through malicious activities or through security policy violations. An IDS gathers and analyzes data from several areas within a computer or a network to identify possible security breaches, which include both intrusions like attacks originating from outside the organization and misuse originating from within the system. An IDS specifically looks for suspicious activities and events that might be the result of a virus, worm or hacker. This is done by looking for known intrusion signatures or attack signatures that characterize different worms or viruses and by tracking general variances which differ from the regular system activity.

Data mining techniques are utilized to supervise and analyze large amount of web data and classify these network data into anomalous and normal information. Data mining techniques can be used for misuse and anomaly intrusion detection. Misuse detection classification is based on known intrusions. Anomaly detection means any substantial differences from the expected behavior are covered as possible attacks. The data mining algorithms used for intrusion detection are classification and clustering. Different classification techniques used in IDS are support vector machines, decision tree, naïve bayes classifier and K-nearest neighbor classifier. Some clustering algorithms are K-means clustering algorithm, K-medoids clustering algorithm.

Support vector machine is a classification algorithm used to classify the attacks in the dataset. Its classification accuracy depends on the parameters used for the kernel function and error cost of the SVM. So that the global optimization algorithm called particle swarm optimization is used to select the optimized parameter values for support vector machine. But a particle swarm optimization suffers from the premature convergence problem which means it converges to suboptimal solutions instead of the global optimal solution. This is because of improper distribution of initial population for the PSO. Therefore the optimal latin hypercube design and classification and regression tree techniques are used to initialize the population.

2. RELATED WORKS

Chung-Jui Tu et al [11] discussed about feature selection using PSO and SVM. The feature selection method based on the number of features investigated for sample classification is needed in order to speed up the processing rate, predictive accuracy, and to avoid incomprehensibility. PSO is used to implement a feature selection and SVM with the one-versus-rest method serve as a fitness function of PSO for the classification problem. This method simplified the feature selection and the total number of parameters needed effectively. Thereby, obtaining higher classification accuracy compared to other feature selection methods. In PSO premature convergence problem was occurring which affects its behavior.

Huang et al [10] suggested a novel PSO-SVM model. This optimization mechanism combined the discrete PSO with the continuous-valued PSO to simultaneously optimize the input feature subset selection and the SVM kernel parameter setting. The hybrid PSO–SVM data mining system was enforced via a distributed architecture using the web service technology to cut back the computational time. To reduce the bottleneck load and to achieve better computational performance, the distributed system should be properly tuned to balance the load among the application server, database server, clients and network traffic. Due to the improper distribution of initial swarm in PSO, it leads into suboptimal solution instead of global optimizing.

Zhan et al [7] extended the PSO to Adaptive Particle Swarm Optimization (APSO). APSO introduces two new parameters to the PSO paradigm. The foremost one is inertia weight to balance the global and the local search capabilities in PSO and another one is control of the acceleration coefficients. It causes the algorithm extremely efficient, providing a substantially improved convergence speed in terms of both numbers of function evaluations and CPU time needed to reach satisfactory results for both unimodal and multimodal functions. APSO will make an impact on the applications of PSO to real world optimization and search problems.

Li et al [6] proposed an improved particle swarm optimization algorithm to train the Fuzzy Support Vector Machine (FSVM) for pattern multi-classification. In FSVM, the truncated polyhedral pyramidal membership functions are presented to resolve unclassifiable regions. PSO introduces stochastic and multipoint searching and find results quickly in high dimensional space. The Improved Linear Particle Swarm Optimization (ILPSO) is introduced for FSVM training for the new particle’s learning method and adaptive variation is considered to develop forth from the local optimum.

Ning et al [5] proposed an improved particle swarm optimization algorithm to optimize penalty parameter and kernel function parameters of SVM. The performance of SVM in network intrusion detection depends on the choice of penalty parameter and kernel function parameter. If the parameters are not correct, then the accuracy rate is not idealistic.

Ahmad et al [4] addressed the false detection rate of intrusion detection approaches. The Genetic Algorithm (GA) is used to explore the genetic principal components that offer a subset of features with the optimal sensitivity and the highest discriminatory power. The SVM is used for classification of attacks. This method provides optimal performance in intrusion detection, which is capable to minimize amount of features and maximize the detection rates.

Rathi et al [2] proposed a technique for intrusion detection using PSO. This methodology uses a GA for feature selection and adaptive mutation for slow convergence of the algorithm. The genetic algorithm is used for solving both constrained and unconstrained optimization problems based on natural selection process that mimics biological evolution. The adaptive mutation is used for omitting the early convergence of PSO. This method is effective and feasible for intrusion detection. PSO has premature convergence problem and it affects the optimized result.

Kuang et al [3] introduced a novel hybrid Kernel Principal Component Analysis (KPCA) and SVM with GA model for intrusion detection. A multi-layer SVM classifier is adopted to gauge whether the action is an attack, KPCA is used as a preprocessor of SVM to reduce the dimension of feature vectors and shorten training time. GA is employed to optimize error cost, kernel parameter of support vector machine. SVM classification model, GA is used to select suitable parameters for SVM classifier, which avoids over-fitting or under-fitting of the SVM model occurring because of the improper determination of the parameters for SVM. The classification accuracy KPCA-SVM model is superior to those of SVM classifiers whose parameters are randomly chosen.

Zhao et al [1] proposed an improved particle swarm optimization algorithm to overcome the premature convergence problem and to improve the optimization capability of PSO. The purpose of the experiments which spreads the initial particles across a design domain and Classification and Regression Tree (CART) which is to identify the promising optimization regions are applied. The improved PSO algorithm is initialized by optimal latin hypercube design. OLHD follows the space filling property in the design space. Thus, successfully enhances the efficiency of the basic PSO.

3. PROPOSED SYSTEM

3.1. Particle Swarm Optimization

The basic PSO is a robust stochastic optimization technique based on the motion and intelligence of swarms. It employs a number of particles that constitute a swarm, moving around in the search space looking for the best solution. Each particle keeps track of its coordinates in the solution space which are affiliated with the best fitness that has achieved so far by that particle. This value is called personal best, pbest. Another best fitness value that is tracked by PSO called global best, gbest. The basic concept of particle swarm optimization lies in accelerating each particle towards its pbest and gbest locations, with a random weighted acceleration at each time step as shown in Figure 1.

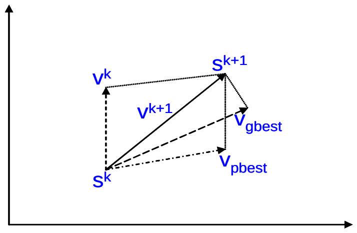

Figure 1: Concept of modification of a searching point by PSO

As given in Figure 1, s^k : current searching point, s^{k+1} : modified searching point, v^k : current velocity, v^{k+1} : velocity based on modified velocity , v_{pbest} : velocity based on pbest, v_{gbest} : velocity based on gbest.

The adjustment of the particle's velocity and position [1] can be mathematically modeled according to the Equation 1 and Equation 2:

$$v_i(k+1) = wv_i(k) + c1.rand1().(p_{best}(i) - s_i(k)) + c2.rand2().(g_{bes\ t} - s_i(k)) \qquad (1)$$

$$Current\ Position\ [k+1] = Current\ Position\ [k] + v\ [k+1] \qquad (2)$$

where, $v_i(k+1)$: velocity of particle at $k+1$th iteration, $v_i(k)$: velocity of particle at kth iteration, w : weighting function, c1 : acceleration factor related to gbest, c2 : acceleration factor related to pbest, rand1(), rand2() : random number between 0 and 1,g_{best} : g_{best} position of swarm, p_{best} : p_{best} position of particle.

3.2. Classification and Regression Tree

Classification and regression tree is a machine-learning method for building prediction models from data. The examples are obtained by recursively partitioning the information space and fitting a simple prediction model within each division. As a consequence, the partitioning can be presented diagrammatically as a decision tree. Classification trees are designed for dependent variables that hold a finite number of unordered values with prediction error measured in terms of misclassification cost. Regression trees are dependent variables that require continuous or ordered discrete values with prediction error typically measured by the squared deviation between the observed and anticipated values.

CART is used to identify promising optimization regions [1] of the input variable space. Established on a series of rules about the attributes of the divisions, the tree is constructed by dividing the information repeatedly. The design variables are the attributes and the classes are feasible values of the target function. The CART technique should follow the purer principle [1] for splitting procedure. Each child node should be purer than its parent. Initially the training dataset is split into two sub-sets by playing along the purer principle. The CART repeats the division of each group separately until the stop criteria reached.

In CART, the decision tree is generated by conforming to three steps:

Step 1: Splitting of design variables

Step 2: Select and execute these splits

Step 3: Stop splitting on a node when predefined termination rules reached

3.3 SVM Classifier

SVM has been applied to intrusion detection schemes for its powers to perform classification and regression. In IDS, SVM is used widely for one class and multi class classification. SVM can also predict different forms that were not employed in training, as it gets generalized after successful cross validation.In classification of two classes, SVM linearly separates attacks from the normal rules. But, if it is not possible to linearly separate the classes, then non linear data are transformed into high dimensional space for classification. In this situation, SVM uses kernel functions for the classification.

In SVM [8], given a training set of instance label pairs (x_i, y_i),i= 1,2,...,m where $x_i \in R^n$ and $y_i \in \{+1, -1\}$, the generalized linear SVM finds an optimal separating hyper plane f(x)= $\langle w \cdot x \rangle$ + b by solving the following optimization problem:

$$\underset{w,b,\xi}{\text{Minimize}} \frac{1}{2}\langle w \cdot w \rangle + C \sum_{i=1}^{m} \xi_i$$

$$\text{Subject to}: \ y_i(\langle w \cdot x_i \rangle + b) + \xi_i - 1 \geq 0, \ \ \xi_i \geq 0$$

(3)

In the Equation (3), *C* is a penalty parameter on the training error and ξ_i is the non-negative slack variables. SVM finds the hyper plane that provides the minimum number of training errors. This optimization model can be solved by introducing the Lagrange multipliers α_i for its dual optimization model. Later on the optimal solution αi* is obtained, the optimal hyper plane parameters w*and b* can be written as Equation (4).

$$\text{sign}\left(\langle w^* \cdot x \rangle + b^*\right)$$

or

$$\text{sign}\left(\sum_{i=1}^{m} y_i\, \alpha_i^* \langle x_i \cdot x \rangle + b^*\right)$$

(4)

The nonlinear SVM maps the training samples from the input space into a higher dimensional feature space via a mapping function Φ. By doing such a mapping, the training samples can be linearly separated by using the linear SVM formulation. The scalar product $\langle \Phi(x_i) \cdot \Phi(x_j) \rangle$ is calculated directly by computing the kernel function k (x_i, x_j) for giving training data in an input space. The Equation (5) is the common kernel function called Radial Basis Function (RBF).

$$k(x_i, x_j) = \exp\left(-\frac{1}{\sigma^2}\|x_i - x_j\|^2\right)$$

or

$$k(x_i, x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right)$$

(5)

For the non-linear classification of data set SVM uses kernel function. It is shown in Equations (6).

$$\text{sign}\left(\sum_{i=1}^{m} y_i \alpha_i^* \langle \Phi(x) \cdot \Phi(x_i) \rangle + b^*\right)$$

or

$$\text{sign}\left(\sum_{i=1}^{m} y_i \alpha_i^* k(x, x_i) + b^*\right)$$

(6)

The performance of SVM mainly depends on its penalty factor and kernel function parameters. The proposed OLCPSO optimization algorithm is applied to select optimized parameters for SVM which in turn produces accurate and robust classification results for intrusion detection.

4. OLCPSO

OLCPSO means Optimal Latin hypercube design and Classification and regression tree techniques for improving basic particle swarm optimization. Optimal latin hypercube design is employed to initialize the population for the particle swarm optimization. The space-filling property is followed to initialize the swarm across the design space. Thus, the Enhanced Stochastic Evolutionary (ESE) algorithm is employed to improve the space-filling qualities among LHD [1]. There are two parts in the initial swarm of OLCPSO. One part is generated by optimal latin hypercube design to cover the design space, while the other part is distributed over the optimal promising regions identified by CART, a classification technique. Figure 2 depicts the process of OLCPSO.

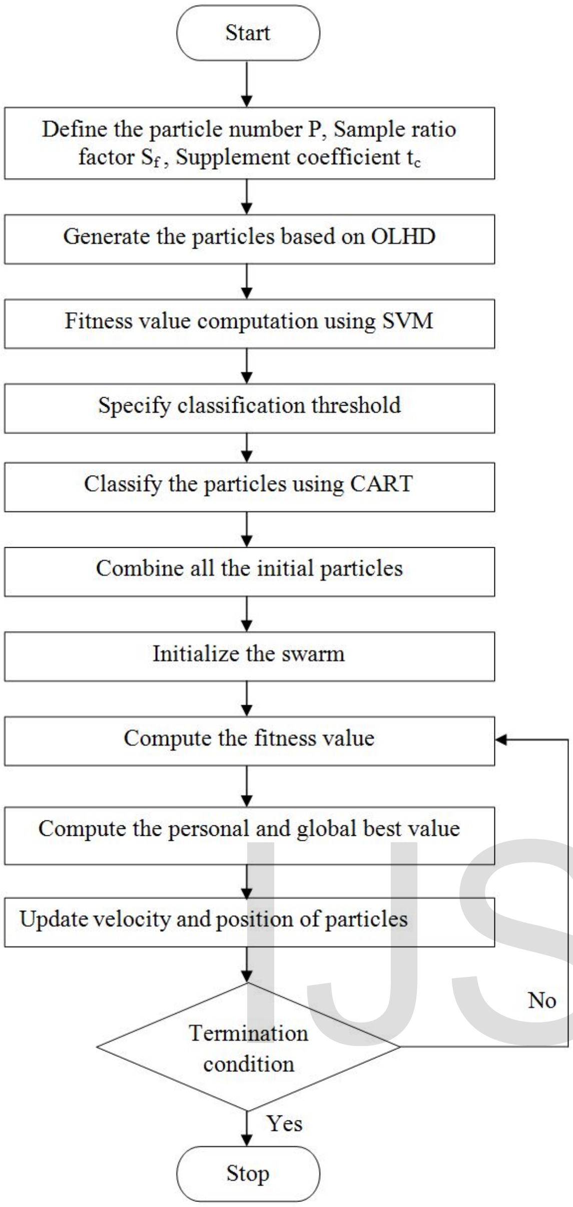
Start
Define the particle number P, Sample ratio factor S_f, Supplement coefficient t_c
Generate the particles based on OLHD
Fitness value computation using SVM
Specify classification threshold
Classify the particles using CART
Combine all the initial particles
Initialize the swarm
Compute the fitness value
Compute the personal and global best value
Update velocity and position of particles
Termination condition
No
Yes
Stop

Figure 2: Flow diagram for OLCPSO

Thus the PSO algorithm is optimized by initializing its population using OLHD and CART techniques. This improved PSO produces optimized results for any application.

5. Experimental Setup and Result Analysis

The database used to examine the proposed OLCPSO-SVM model for intrusion detection is Knowledge Discovery in Database (KDD) Cup data which is published by the DARPA in 1999. It admits a broad variety of intrusions like probes, denial-of-services attacks. The dataset has 41 features and 24 attack types, but treats all of them as an attack group.

The performance of the OLCPSO-SVM model is examined by considering accuracy. From the Table 1, it is known that the proposed OLCPSO-SVM yields 94.09% classification accuracy which is larger than the existing support vector machine model combining kernel principal component analysis with genetic algorithm (KPCA-GA-SVM) [3]. Figure 3 shows the comparison graph of proposed and existing systems.

Table 1 Comparison of KPCA-GA-SVM and OLCPSO-SVM Models

Algorithm	**Classification Accuracy rate (%)**
KPCA-GA-SVM	92.065
OLCPSO-SVM	94.09

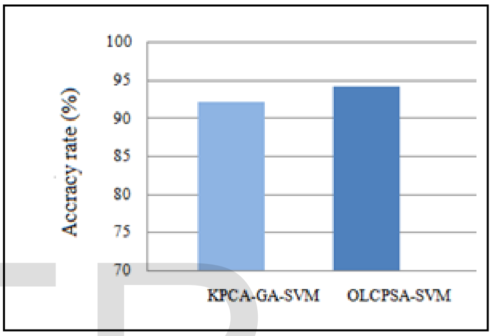

Figure 3: Comparison graph of KPCA-GA-SVM and OLCPCA-SVM

Therefore, increase in measure of accuracy is because of optimized parameters of SVM and OLCPSO-SVM system shows better performance than KPCA-GA-SVM in intrusion detection.

6. Conclusion

Optimal LHD and CART are used to bring forth the initial population for the PSO. So the initial population of OLCPSO is generated in design space and supplement region using optimal latin hybercube design and CART respectively. It improves the optimization performance of PSO. OLCPSO is applied to optimize the penalty parameter and kernel function parameter of SVM. The performance of SVM in network intrusion detection is increased. The proposed OLCPSO-SVM model addresses the over fitting problem. Thus, the OLCPSO-SVM model is applied to intrusion detection dataset that yields high detection rates for the intrusion detection system. In future, optimization algorithm can be applied to optimize support vector machine model feature selection for the intrusion detection system.

REFERENCES

[1] ‘Improved Particle Swarm Optimization Algorithm Using Design Of Experiment And Data Mining Techniques’, Zhao Liu, Ping Zhu,Wei Chen, Journal on Structural and Multidisciplinary Optimization ,Volume No.52, Issue No.8, pp.1-14, 2015.

[2] ‘Network Intrusion Detection Using PSO Based on Adaptive Mutation and Genetic Algorithm’, Bharat Rathi, Dattatray V. Jadhav, International Journal of Scientific & Engineering Research, Volume No.5, Issue Issue No.8, pp.142-144, 2014.

[3] ‘A novel hybrid KPCA and SVM with GA model for intrusion detection’, Fangjun Kuang, Weihong Xu, Siyang Zhang, Journal of Applied Soft Computing, Volume No.18, Issue No.8, pp.178-184,2014.

[4] ‘Enhancing SVM performance in intrusion detection using optimal feature subset selection based on genetic principal components’, Iftikhar Ahmad, Muhammad Hussian and Abdullah Alghadi, Journal of Neural Computing and Applications, Volume No.24, Issue No.7, pp.1671-1682, 2013.

[5] ‘Intrusion Detection Research Based on Improved PSO and SVM’, Liu Ning, Zhao Jianhua, Proceedings of International Conference on Automatic Control and Artificial Intelligence, pp.263-1264, 2012.

[6] ‘Improved particle swarm optimization algorithm for fuzzy multi-class SVM’, Ying Li, Bendu Bai and Yanning Zhang, Journal of Systems Engineering and Electronics, Volume No.21, Issue No.3, pp.509-513, 2010.

[7] ‘Adaptive Particle Swarm Optimization ’, Zhi Hui Zhan, Jun Z hang, Yun Li and Henry Shu Hung Chung, IEEE Transactions on Systems, Man and Cybernetics, Volume No.39, Issue No.6, pp.1362-1381, 2009.

[8] ‘Intrusion Detection Method Based on Classify Support Vector Machine’, Meijuan Gao, Jingwen Tian and Mingping Xia, Proceedings of Second International Conference on Intelligent Computation Technology and Automation, pp.391-394, 2009.

[9] ‘Effective value of decision tree with KDD 99intrusion detection datasets for intrusion detection system’, J.H. Lee, J.H. Lee, S.G. Sohn, et al., Proceedings of International Conference on Advanced Communication Technology, pp. 1170–1175, 2008.

[10] ‘A distributed PSO–SVM hybrid system with feature selection and parameter optimization’, Cheng-Lung Huang, Jian-Fan Dun, Journal of Applied Soft Computing, Volume No.8, Issue No.3, pp.1381-1391, 2007.

[11] ‘Feature Selection using PSO-SVM’,Chung-Jui Tu, Li-Yeh Chuang, Jun-Yang Chang, and Cheng-Hong Yang, IAENG International Journal of Computer Science, Volume No.33, Issue No.1, pp.1-18,2007.

[12] ‘Intrusion detection model based on particle swarm optimization and support vector machine,’S. Srinoy, Proceedings of IEEE Symposium on Com-putational Intelligence in Security and Defense Applications, pp. 186-192, 2007.

[13] ‘ A hierarchical particle swarm optimizer and its adaptive variant’, Janson and Middendorf, IEEE Transactions on Systems, Man, and Cybernatics -Part B: Cybernatics, Volume No.35, Issue No.6, pp.1272-1282, 2005.