International Journal of Security and Its Applications
Vol.8, No.3 (2014), pp. 249-260
http://dx.doi.org/10.14257/ijsia.2014.8.3.26

Research on Intrusion Detection System Based on Clustering Fuzzy Support Vector Machine

Zhai Jinbiao[1]

State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, BeiHang University

jinbiaozhai@gmail.com

Abstract

Introducing the artificial intelligence learning algorithm to solve the problem of network security is a focus of current research. We introduce the clustering algorithm into artificial intelligence learning algorithm and apply Fuzzy Support Machines to the intrusion detection. We put forward a method which is based on Fuzzy Support Machines. Then, we chose an appropriate RBF kernel function according to the characteristic of intrusion detection. And we get the intrusion detection algorithm based on Fuzzy Support Machines. The algorithm in this paper reduces the training time and improves the efficiency of the algorithm. Experimental results show that this method improves the fuzzy support vector machine training efficiency, and it is also very effective in intrusion detection. The first part of this paper is the introduction of the related problem. The second part is the concept of Fuzzy Support Vector Machine. The third part is the choice of the clustering center. The fourth part is the process of intrusion detection algorithm. The final part is the experiment.

Keywords: *Intrusion Detection, Fuzzy Support Vector Machines, Network security*

1. Introduction

Support Vector Machine (SVM) is the core content in the statistical learning theory (SLT). It is based on the VC dimension theory and structural risk minimization principle. Statistical learning theory overcomes the curse of dimension in the traditional machine learnin g and local minimum problem [1, 2]. We have proven that the general ability of SVM is the best in dealing with the problem of small samples size. But SVM has some performances that the traditional machine learning methods can't match though it is not the best classification accuracy at times. This theory has been widely applied in data classification and regressive problems such as handwritten numeral recognition, object recognition, speech recognition and spatial data analysis and so on based on the above reason.

In SVM theory, the original input space is mapped to a high dimensional space by using a kernel function. High dimensional space calls feature space. In this space, an optimal hyper plane will maximize the generalization ability of the classifier. The optimal hyper plane can be generated by a small number of data points and these points are called Support Vector (SV). Therefore, SVM can also get a good classification results without knowing the original problem domain knowledge. This characteristic is one of the advantages for Support Vector Machine [1~3]. At present, there are some methods which can solve multi-class problems. For example, the DAGSVM method [4], the multi-class SVM classifier method Weston and Watkins proposed [5], the 1-v-1 method and a large interval of DAG multi-class SVM algorithm Platt proposed [7,8].

ISSN: 1738-9976 IJSIA

Intrusion is defined as a behavior that any behaviors attempt try to endanger the computer resources integrity, confidentiality or availability. Research on intrusion detection can be traced to 20 centuries 80 years. After the rapid development of the Internet, we begin to pay attention to it.

W. Lee proposed and realized multilevel IDS in the CDF (common intrusion detection framework). And in 1999, they discussed the audit data processing by using data mining technology [9]. In 2000, Ghosh used a neural network to extract feature and classification [10]. Application and method of SVM in intrusion detection have been developed in recent years. In 2002, Mukkamala and the other scholars used SVM technology to achieve the intrusion detection [11]. This method is compared with the neural network method and it gets the better results. B.V.Nguyen and E.Eskin proposed anomaly detection which is based on one class SVM [12].

At present the fuzzy theory is applied to intrusion detection in the following two representative: the experimental results of Sung-BaeCho [13]. It adopts fuzzy logic, HMM model and the self-organizing feature map neural network together. Its effect is ideal. The test result of IDS false rate is 1.18%. Fuzzy Support Vector Machine model is proposed by Chun-Fu Lin and Sheng-De-Wang in literature [14]. Its aim is to use the SVM to apply practical problems. Liu Guangli and the other scholars put forward the concept of uncertainty Support Vector Machine [15]. They prove that the Fuzzy Support Vector Machine is a special case of the uncertainty Support Vector Machine. And they introduce it into China food security warning system. This method solves the problem of determining the degree of alarm history effectively.

Some scholars put forward a Fuzzy Support Vector Machine intrusion detection algorithm based on clustering. We propose a method that can prune the training data and reduce the number of edge points which are away from the clustering classification effectively. This method keeps more samples near the classification surface in order to close to the boundary of the discriminant clustering center set. It is an effective training sample to train the Fuzzy Support Vector Machine set. And it also reduces the training time and improves the efficiency of the algorithm.

2. Fuzzy Support Vector Machine

Intrusion Detection System (IDS) is a kind of active network security protection system. It is a new generation of security technology after the traditional security technology of the data encryption and firewall etc. IDS provides for internal attack, exterior attack and operation real-time protection. It responds the intrusion before the host or the network system is attacked. At present, there are many ways of intrusion detection. It mainly divides into anomaly detection and misuse detection. Intrusion detection technology is a hot topic now. However, with the diversity of intrusion techniques, traditional intrusion detection technology can not meet the current requirements of network security. The combination between intelligent technology and IDS has become a hot topic of current research. In recent years, the main technologies which the field of intrusion detection uses are Bayesian inference, artificial neural network, expert system, computer immunology and data mining etc. But now, the popular intrusion detection system is false positives, false, poor real-time characteristics and longer training time. In recent years, some scholars have introduced the Support Vector Machine into the intrusion detection and achieved good results. Intrusion Detection Technology which is based on Support Vector Machine has become a hot topic in the field of intrusion detection. In this paper, we combine the clustering algorithm and fuzzy support vector machine method in order to improve the learning speed. We make the clustering center

set as the training data of SVM algorithm. The aim of this measure is to reduce the training time and improve the efficiency of the algorithm.

Support Vector Machine is short for SVM. SVM develops from the optimal linear classification hyper plane. The basic idea can be illustrated by the two dimensional from figure 1.In figure 1, there are two different spots which stand for two samples. H is the classification hyper plane. These two different points are the lines which represent the samples that go through the nearest hyper plane and parallel to the hyper plane. The distance between them is called the margin. The optimal hyper plane is a hyper plane which can not only separate two types correctly (training error rate is 0)but also make classification interval.

SVM is proposed by Vapnik. It is a kind model which is based on statistical learning theory. In the nonlinear case, the form of the optimal classification function $f(x)$ is as following:

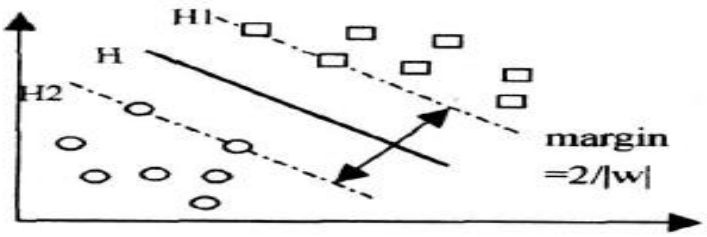

Figure 1. The Optimal Hyperplane

$$f(x) = \mathrm{sgn}(wx + b) = \mathrm{sgn}(\sum_{i=1}^{l} (a^* = a_i) k(x_i, x) + b) \quad (1)$$

Among them, w is weight vector. b is offset value. $k(x_i, x)$ kernel function. The common kernel functions have:

$$k(x, y) = x^T y \text{ (the dot product kernel function)} \quad (2)$$

$$k(x, y) = (x^T y + c)^d \text{ (polynomial kernel function)} \quad (3)$$

$$k(x, y) = \exp(\frac{\| x - y \|^2}{2\sigma^2}) \text{ (the radial kernel function)} \quad (4)$$

SVM is sensitivity to the noise and the outlier. But its noise tolerance is poor. Therefore, how to improve the noise tolerance of the vector machine has become a very important topic. In order to improve the anti-noise ability of the data for a vector machine, we can introduce a fuzzy factor $\mu_i, 0 \le \mu_i \le 1$.We define it as $\tau_i = \frac{1}{\mu_i}$.On the basic of the above content, we can transfer the support vector optimization problem into the following optimization problem:

$$\min T(\omega \zeta^{(*)}, \varepsilon) = \frac{1}{2} \| \omega \|^2 + \frac{C}{l} \sum_{i=1}^{l} [\gamma \varepsilon + \tau_i (\zeta_i + \zeta_i^*)] \quad (5)$$

$$s.t. \quad \omega x_i + b - y_i \le \varepsilon + \zeta_i \quad (6)$$

$$y_i - (\omega x_i + b) \le \varepsilon + \zeta_i^* \qquad \varepsilon, \zeta_i^* \ge 0$$

In the formula, C is the penalty coefficient. It is used to control the model complexity and the training error. $C > 0$. ε is the training precision parameter. γ is used to control the size of the pipe ε, $0 < \gamma < 1$. $\zeta_i^{(*)}$ is the slack variables. $\zeta_i^{(*)}$ is used to correct the correction of positive constraints. τ_i is the inverse of fuzzy factor. When τ_i is smaller, the effect of the corresponding sample on the optimization problem is smaller. If we solve the optimization problem of the type (5) and (6), we can introduce Lagrange function.

$$\begin{aligned} L(\omega, \zeta^{(*)}, \beta, \varepsilon, \eta^{(*)}) &= \frac{1}{2} \| \omega \|^2 + \frac{C}{l} \sum_{i=1}^{l} [\gamma\varepsilon + \tau_i(\zeta_i + \zeta_i^*)] \\ &- \beta\varepsilon - \sum_{i=1}^{l} (\eta_i \zeta_i + \eta_i^* \zeta_i^*) - \sum_{i=1}^{l} \alpha_i (\varepsilon + \zeta_i + y_i - \omega x_i - b_i) \\ &- \sum_{i=1}^{l} \alpha_i^* (\varepsilon + \zeta_i^* - y_i + \omega x_i + b_i) \end{aligned} \tag{7}$$

By KKT conditions, the function type (7) is 0 respect to the variables. Among them, $\alpha_i^{(*)}$ is the multiplier of Lagrange.

$$\frac{\partial L}{\partial \omega} = 0 \Rightarrow \sum_{i=1}^{L} (\alpha_i^* - \alpha_i) x_i \tag{8}$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{L} (\alpha_i^* - \alpha_i) \tag{9}$$

$$\frac{\partial L}{\partial \varepsilon} = 0 \Rightarrow C\gamma - \beta - \sum_{i=1}^{L} (\alpha_i^* + \alpha_i) = 0 \tag{10}$$

$$\frac{\partial L}{\partial \zeta^*} = 0 \Rightarrow \frac{C}{l} \tau_i - y_i - \alpha_i^{(*)} \tau_i = 0 \tag{11}$$

We take (8),(9),(10),and (11) into the optimization problems of the Lagrange function (7).We need to solve the maximum value about $\alpha_i^{(*)}$,then we can obtain the dual optimization problem of the fuzzy support vector machine.

$$\min W(\alpha^{(*)}) = \frac{1}{2} \sum_{i,j=1}^{l} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) k(x_i, x_j) - \sum_{i=1}^{L} (\alpha_i^* - a_i) y_i \tag{12}$$

$$s.t. \quad \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) = 0 \tag{13}$$

$$0 \le \alpha_i^{(*)} \le \frac{C}{l} \tau_i$$

$$\sum_{i=1}^{l} (\alpha_i^* + \alpha_i) \le C\gamma$$

From (12) and (13),we can calculate the optimal value α_i and α_i^*.Then we can get the classification function of FSVM:

$$f(x) = \operatorname{sgn}(\sum_{i=1}^{l} (\alpha_i^* - \alpha_i) k(x_i, x_j) + b) \tag{14}$$

3. The Selection of the Efficient Clustering Center Set

Support Vector Machine is based on statistical learning theory. SVM can solve learning problems better for the small samples because we use the structural risk minimization to replace the empirical risk minimization principle. SVM adopts the kernel functions. It can convert the non-linear problem into a linear problem so that it reduces the complexity of the algorithm. SVM becomes the research topic after neural network because of the more complete theoretical basis and better learning performance. But the method of SVM about the classification problems is sensitive to the noise. Lin and the other people introduce the concept of fuzzy membership into the SVM classification. They also propose the Fuzzy Support Vector Machine. However, there are some factors influencing the complexity of the fuzzy SVM classifier. For example, the size of the training sample, the dimension of input samples, the number of support vectors and the distribution if support vectors and so on. The low speed of training algorithm is one of the bottlenecks to influence the wide application of SVM.

Clustering analysis is a non-supervised learning method. Its main function is spatially detailed data mining. The goal is the similar things with similar properties. Clustering divides physical or abstract data object into different groups of classifications according to the similarity of objects. Clusters which are generated by the class or cluster are a set of data objects. A lot of simple classification algorithms which are applied to the pattern recognition problems are based on k-nearest neighbor method. The method is very important because it can be used as learning based on memory, learning based on instances and learning based on requirements in the field of artificial intelligence. The k-nearest neighbor method stores all the training samples. This method needs not to require structural model before accepting the new samples which are to be classified. And it needs to establish a new classification until the new sample needs. The k-nearest neighbor method is based on learning by analogy. The training sample has k value dimension attribute description. A sample represents the K dimensional space. All of the training samples are stored in the K dimensional pattern space. Given an unknown sample, k-nearest neighbor search pattern space can find the K training samples closest to the unknown samples. These k samples are k neighbor unknown samples. Based on the principle, we can use k-nearest neighbor method to prune the intrusion detection data set. We use the clustering center set as the training samples of SVM to train. And this method can reduce the training time of the algorithm.

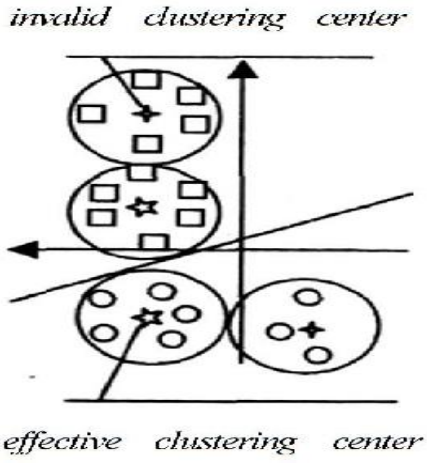

Figure 2. Schematic Diagram of Clustering Center Set Selection Algorithm

The method uses the k-nearest neighbor to prune effectively for the data set. It adopts the clustering center closest to the decision boundary as the set of support vectors of the training samples. If the two categories respectively p_1 and p_2, the two centers are represented as a set:

$$C_1 = \{C_{11}, C_{12}, \cdots, C_{1n}\}$$

$$C_2 = \{C_{21}, C_{22}, \cdots, C_{2n}\}$$

$$(C_{1i}, C_{2j}) = \min \| C_{1i} - C_{2j} \|^2 \qquad (1 \le i \le n, 1 \le j \le m)$$

They are sets PC_1 and PC_2 as the first element of the effective clustering center set. If the average of the distance between a cluster center PC_1 and each element is more than the average of the distance between a cluster center PC_2 and each element, then we select the cluster center to join PC_2 .Or we select the cluster center to join PC_1 .We assume that $n_1 = | PC_1 |, m_1 = | PC_2 |, k \in \{1, 2\}$,

$$d = \frac{\sum_{i=1}^{n_1} \| C_{kl} - C_{1i} \|^2}{n_1} - \frac{\sum_{j=1}^{m_1} \| C_{kl} - C_{2j} \|^2}{m_1} .$$

If $d \ge 0$ then $k = 1$,we make C_{kl} join to PC_2 .If $d < 0$ then $k = 2$,we make C_{kl} join to PC_1 .We repeat the process until all the clustering center do not satisfy the condition of formula 1.The elements of PC_1 and PC_2 are efficient clustering centers near the boundary. We make the selected effective clustering center set as effective training samples sets of two categories. Then we train the fuzzy support vector machine.

4. Vector Machine Intrusion Detection Algorithm based on Clustering Fuzzy Support

Firstly, Fuzzy Support Vector Machine algorithm based on clustering prunes sets effectively. According to the differences between each sample and the nearest class, we decide how to choice. The nearest neighbor classification is based on learning by analogy and it gives an unknown sample. The k-nearest neighbor classification method searches the model space and finds the K training samples of unknown samples. The K samples is K neighbor unknown samples. For each sample point, we keep this point if the point and the nearest neighbor belong to the same class. If the point and the nearest neighbor belong to heterogeneous, we delete the point.

In order to improve the learning speed of Support Vector Machine, we combine the Support Vector Machine and clustering algorithm. We use the k-means clustering algorithm to cluster the training data in order to get the clustering center set of each clustering. Then, we make the clustering center set as the training data. Each clustering corresponds to a discriminant function of a Support Vector Machine. The learning process of training samples is showed in figure 3. When we test on a data set, we should carry on the localization to the test data. We use the discriminant function of the clustering class to determine the type of the data.

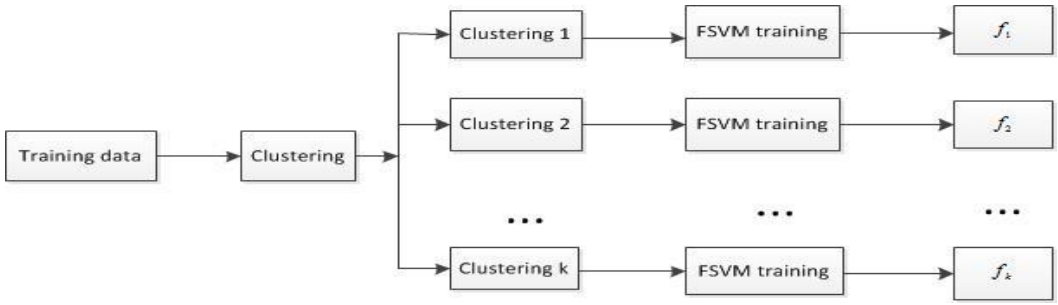

Figure 3. Clustering Fuzzy Support Vector Machine Learning Algorithm based on Process Map

The step description of Support Vector Machine algorithm based on clustering is as following: A. We use k-nearest neighbor method to prune data preprocessing for the training sample sets. B. We make the training data set by using the K-means clustering algorithm for clustering. C. We get each clustering center set. D. We use the clustering center set as training samples to generate the discriminant functions. We generate a FSVM discriminant function in each clustering. E. According to the test data, we locate the clustering category of the test data. F. According to the discriminant function of locating clustering, we determine the category of the test data. The chart of the algorithm is following:

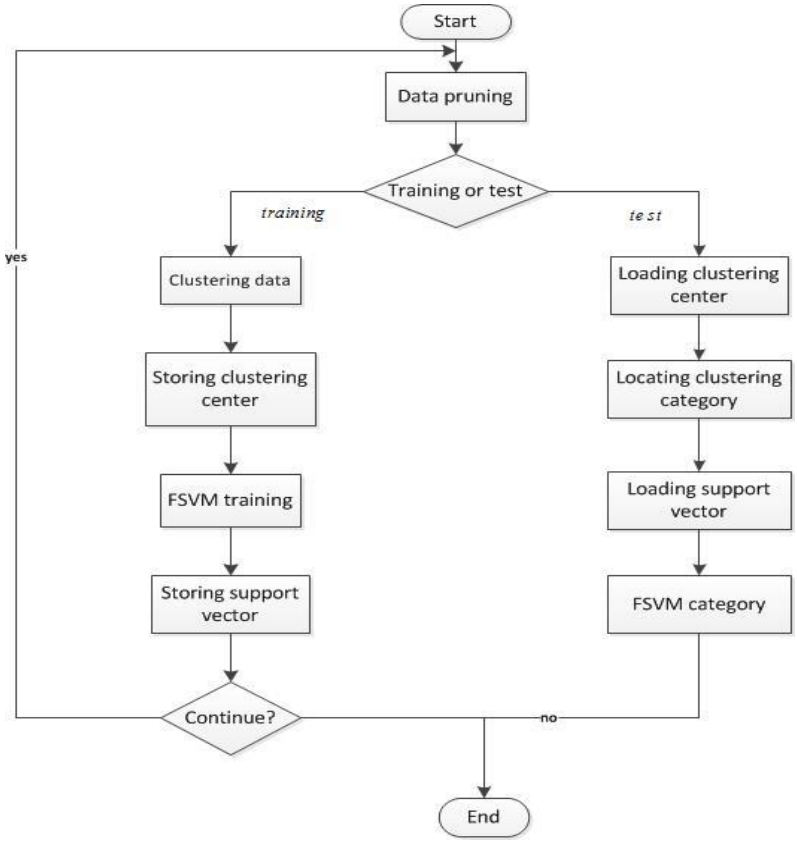

Figure 4. Flow Chart of the Algorithm

5. Experiment

5.1. Numerical Experiments

Firstly we select seven data sets in the UCI machine learning database. They are Iris, glass, soy, vowel, blood cell and Thyroid. In order to verify the applicability of FVSM about the noise data we use NDC data generator to generate a noisy data set FTRAIN. FTRAIN contains 1200 samples. The dimension of each sample is eight. The number of class samples is 3.The sample data obey a normal distribution. On this basis, we add 80 noise data (mean is 0 and spectral noise density is a normal number).These data will be divided into two groups in the experiment. Table 1 describes simply the basic attribute of the data. We compare the FVSM algorithm proposed in this paper with the multi-class VSM algorithm. In table 1,#pts is the number of data point and #att is the number of attributes. In the experiment we take $s_i = {}^{y_i}\!/_{k}$ .k is the number of categories. $y_i = 1, 2, \cdots k$ is class label.

In order to make the results more convincing, we select different C value for each algorithm. And we use two kernel functions: Polynomial kernel function and Gauss kernel function. Each algorithm in the implementation process uses several different values to compare such as the number d of polynomial kernel function and Gauss kernel function span coefficient .We adopt 10 fold cross validation method to estimate the accuracy of classifiers. Table 2 and Table 3 give the correct average rate of classification. The two kernel functions are:

$$K(x, y) = [(x \cdot y) + 1]^d \text{ (polynomial kernel function)}$$

$$K(x, y) = \exp(-g \| x - y \|^2) \text{ (Gauss kernel function)}$$

Table 1. The Data Set Used in the Experiments

Data set	#pts	#att	The number of categories
iris	150	3	4
wine	178	13	3
glass	214	9	7
soy	289	208	17
vowel	528	10	11
Blood cell	3097	13	12
thyroid	3722	21	3

According to the analysis of the experimental results, we can see that the performance of FSVM algorithm in the selected data sets is better than others. The classification results in glass, soy and vowel are better than others. But in iris and wine, the classification results are worse than 1-v-1.In blood cell, the collective classification rate is higher than other algorithms. But in thyroid the correct classification, rate is lower than the 1-v-1algorithm and higher than other algorithms. In table 2 and table 3, data means the data set. Kernel means the kernel function. Parm means the parameters. J&C means the multi class SVM algorithm which Weston proposed.

Table 2. The Experimental Results on the First Data Sets

Data	Kernel	Parm	1-v-r(%)	1-v-1(%)	J&C(%)	FSVM(%)
iris	Poly	4.0	96.00	96.29	96.24	96.34
		5.0	96.89	96.94	96.86	96.86
		6.0	96.87	96.89	96.32	96.33
	RBF	0.1	97.02	97.02	97.02	97.08
		1.0	96.81	96.89	96.56	96.65
wine	Poly	3.0	97.46	97.56	97.34	97.78
		4.0	97.57	97.53	97.57	97.65
	RBF	3.0	97.76	97.77	97.76	97.78
		0.2	97.56	97.64	97.70	97.65
glass	Poly	3.0	82.12	82.25	81.25	83.30
		5.0	82.12	82.45	82.33	85.30
	RBF	0.3	81.56	83.35	82.46	85.50
		1.0	82.56	83.75	82.56	84.50
soy	Poly	0.5	96.66	97.35	97.56	97.00
		2.0	97.66	97.35	97.26	97.50
	RBF	4.0	95.66	96.35	97.22	97.50
		5.0	95.68	96.33	97.22	97.40
vowel	Poly	0.2	96.66	97.35	97.46	97.50
		2.0	96.66	97.35	97.46	97.50
	RBF	5.0	96.65	97.33	97.45	97.55
		6.0	96.65	97.36	97.45	97.55

Table 3. The Experimental Results on the Second Data Sets of Blood Cell and Thyroid

Data	Kernel	Parm	1-v-r(%)	1-v-1(%)	J&C(%)	FSVM(%)
Blood cell	Poly	4	90.25	92.10	90.35	91.33
		5	91.03	91.90	90.20	92.10
	RBF	6	91.25	91.58	91.24	92.12
		10	92.52	91.58	91.24	92.12
Thyroid	Poly	4	92.27	96.56	96.62	93.54
	RBF	10	92.30	95.75	95.16	95.46
FTRAIN	RBF	8	75.05	76.36	73.23	80.27
		10	76.30	76.35	75.00	81.02

The experimental results show the classification accuracy of FSVM is generally improved. The algorithm is very effective in processing noisy data. And the classification accuracy of the algorithm is significantly higher than the 1-v1 algorithm and the J&C algorithm according to the statistical analysis of the experimental results.

5.2. Application in Intrusion Detection

Currently computer network security is one of the most important research fields in computer science and technology. Computer network security is a comprehensive subject. And this subject includes computer science, network technology, communication technology, cryptography technology, applied mathematics and information theory. Intrusion Detection

Technology is the core content of dynamic security technology. It takes the invasion of active surveillance, monitoring and identification of ongoing attempt. It has successful invaders or intrusive behavior rather than passively defense strategy. Intrusion Detection can also prevent the intrusion events. It can prevent the losses caused by the invasion. It can also prevent the system or data recovery and treatment for intrusion events. In addition it provides information services and evidence.

In essence, the problem of intrusion detection is a problem of pattern recognition and pattern classification problems. SVM is an effective tool to solve these problems. In the paper, this data is available in the invasion model data generation program to generate test sets. The data set has a strong practical. We write an intrusion pattern data creation program to generate an intrusion data set automatically. This program is based on the analysis of all kinds of network intrusion attack and the in-depth understanding of intrusion. The data contain not only various types of attack patterns (a total of four types) but also a certain amount of noise data. The data set has a very strong representation.

Data set contains 4000 intrusion data. Each data is in the 24 dimension vector. Data set contains 100 noise data. The types of attacks are a total of four. The number of data set is 4 categories. These four categories are Dos(denial of service attack), R2L(remote access permissions), U2R(all right ascension) and Probe(Port scanning). The aim of the experiment is to detect the rate and the accuracy of various classifiers data from various attacks.

Detection Rate=detected attack sample number/the normal number of samples. False Positive=normal samples were false alarm for abnormal sample number/the normal number of samples. In the experiment the value of s_i is: the normal data $s_i = 1$;Class Dos attack data: $s_i = 0.8$;Class Probe attack data: $s_i = 0.6$;Class U2R attack data: $s_i = 0.4$;class R2L attack data: $s_i = 0.2$ .The experimental results are listed in table 4.

Table 4. In Intrusion Detection Data Experiments

algorithm	Attack classes			
	DoS	Probe	U2R	R2L
Naive bayes	DR=91% FP=8%	DR=35% FP=6%	DR=23% FP=3%	DR=21% FP=3%
1-v-r	DR=87% FP=7%	DR=52% FP=5%	DR=35% FP=4%	DR=17% FP=2%
1-v-1	DR=93% FP=8%	DR=46% FP=6%	DR=33% FP=5%	DR=12% FP=4%
J&C	DR=91% FP=10%	DR=45% FP=6%	DR=35% FP=5%	DR=33% FP=2%
FSVM	DR=93% FP=10%	DR=91% FP=7%	DR=31% FP=4%	DR=91% FP=8%

The experimental results show that this method has the same detection rate of 1-v-1algorithm in the Dos data. The false detection rate is slightly higher than 1-v-1 algorithm. In the U2R data the detection rate is the lowest. False detection rate is not much difference between with other methods. And in the other case this method is almost the best (the error detection in the Probe is slightly higher but it is higher than other methods).

5.3. Intrusion Detection Experiment

We use the data: KDD CPU 99 for Intrusion Detection Experiment. The data set includes nearly 5000000 simulated attack records. The experimental data we used is about 5% or so subset. Each record description consists7 classification attributes and 34 numerical attributes. We remove the normal data and the data which are not belong to a denial of service attack. So the data only include back, land, Neptune, pod, smurf, teardrop and normal. Data preprocessing is: data is divided into $y = 1$ (classification label for Norma) and $y = -1$ (classification labels for other) two. At last we use the method of minimum value and maximum value to standardize the data. The results are shown at Table.5.

Table 5. Intrusion Detection Algorithm Performance based on FSVM

Test data	Clusterin.g(s)	SVM training (s)	Locating(ms)	Test(ms)	False alarm rate%	False negative rate%
2000	127	18.473	0.07581	0.08841	0.2213	0.1034
4000	502	45.649	0.14630	0.9941	0.1744	0.0103
6000	1750	60.570	0.19786	0.08822	0.1374	0.0091
8000	2140	73.825	0.23412	0.08478	0.1294	0.0053
10000	3101	98.502	0.3100	0.07850	0.0522	0.0012
14000	8402	153.241	0.0743	0.07413	0.0213	0.0004

And then, we compare the traditional SVM and the FSVM. The results are showed as Table.6.

Table 6. Results Compare between Traditional SVM and the FSVM

The sample proportion	FSVM detection rate%	SVM detection rate%
10%	79.0	78.3
20%	85.5	83.0
30%	86.4	83.2
40%	91.8	83.6
50%	92	89.6

From the above experimental results, we can see that the accuracy and the prediction of time efficiency have a marked improvement. The improving algorithm is effective especially for large-scale data set. The improving algorithm is feasible.

6. Conclusion

In this paper, we put forward a kind of FSVM algorithm to solve the existing problems that SVM and related fields meet. It is the fuzzy multi class Support Vector Machine algorithm for intrusion detection algorithm. The method can prune the training data. And it can also reduce the number of edge points effectively which are away from the clustering classification. In order to close to the boundary of the discriminant for clustering center set, this method keeps more samples near the classification surface. Experimental results show the effectiveness of the method. This method improves the generalization ability of fuzzy support vector greatly. And it also improves the widely application range of fuzzy support vector machine.

References

[1] J. C. Burges, "A tutorial on support vector machines for pattern recognition", Data Mining and Knowledge Discovery, vol. 2, no. 2, **(1998)**, pp. 121~1672
[2] V. Vapnik, "Statistical Learning Theory", New York: Springer Verlag, **(1998)**.
[3] C. Corts and V. Vapnik, "Support vector networks", Machine Learning, vol. 20, no. 3, **(1995)**, pp. 273-297.
[4] C. W. Hsu and C. J. Lin, "A comparison of methods for multi-class support vector machines", IEEE Transactions on Neural Networks, vol. 13, no. 2, **(2002)**, pp. 415-425.
[5] J. Weston and C. Watkins, "Multi-class support vector machines", Department of Computer Science, Royal Holloway University of London Technical Report, SD TR 98 04, **(19986)**.
[6] K. Ulrich, "Pairwise classification and support vector machines", Schkopf B., Burges C.J.C., Smola A.J. eds. Advances in Kernel Methods—Support Vector Learning, Cambridge, MA: MIT Press, **(1998)**, pp. 255-268.
[7] J. C. Platt, N. Cristianini and T. J. Shawe, "Large margin DAG's for multiclass classification", Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, vol. 12, **(2000)**, pp. 547-553.
[8] L. Kun Lun, H. Hou Kuan and T. Sheng Feng, "A novel multi-class SVM classifier based on DDAG", Proceedings of IEEE ICMLC'02, Beijing, China, vol. 3, **(2002)**, pp. 1203-1207.
[9] W. Lee and S. J. Stolfo, "A data mining framework for building intrusion detection model", Proceedings of the 1999 IEEE Symposium on Security and Privacy, Oakland, CA, IEEE Computer Society Press, **(1999)**, pp. 120-132.
[10] A. K. Ghosh, C. Michael and M. Schatz, "A real-time intrusion detection system based on learning program behavior", Recent Advances in Intrusion Detection (RAID 2000), Spinger-Verlag, **(2000)**, pp. 93-109.
[11] S. Mukkamala, G. L Janoski and A. H. Sung, "Intrusion Detection Using Neural Networks and Support Vector Machines", Proceedings of IEEE International Joint Conference on Neural Networks, **(2002)**, pp. 87-90.
[12] E. Eskin, A. Arnold, M. Prerau, L. Portnoy and S. Stolfo, "A Geometric Framework for Unsupervised Anomaly Detection: Detecting Intrusions in Unlabeled Data Applications of Data Mining in Computer Security", Kluwer, **(2002)**, pp. 537-547.
[13] S.-B. Cho, Member IEEE "Incoporating Soft Computing Techniques Into a Probalistic Intrusion Detection System", IEEE, **(2002)**.
[14] C.-F. Lin and S.-D. Wang, "Fuzzy Support Vector Machines", IEEE, **(2002)**.
[15] L. Guang-li, Z. Ai-li and D. Nai-yang, "Uncertainty Support Vector Classification for Early-Warning", vol. 8.11, no. 4, **(2003)**, pp. 58-61.

Author

Zhai Jinbiao, borns in 1983. He received his bachelor's degree from Beihang University. Now, he is a PhD in Beihang University. His main research interests include software reuse, distributed workflow, computer security.

