The 4th IEEE International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems
20-21 September, 2018, Lviv, Ukraine

Research on Network Intrusion Detection Based on Support Vector Machine Optimized with Pigeon-inspired Optimization Algorithm

Yiheng Sun[1], Zhiwei Ye[1], Chunzhi Wang[1], Lingyu Yan[1]， Ruoxi Wang[2]
[1]School of computer Science, 430068,yihengsun01@163.com,Hubei university of Technology, Wuhan, China
[2]Wuhan FiberHome Technical Services Co. Ltd. ,430074, Wuhan, China

Abstract—As an important technology in the field of network and information security, intrusion detection plays an important role in the information security protection system, and support vector machine (SVM) is one of the most successful methods. However， the performance of SVM is affected by its parameters. In order to improve the network intrusion detection accuracy, pigeon-inspired optimization (PIO) is introduced into intrusion detection to optimize the SVM parameters (PIO-SVM). PIO-SVM uses network intrusion detection data as the inputs of SVM, and SVM parameters are optimized by pigeon individual in PIO, the network intrusion detection accuracy is used as PIO target function, and then through mutual cooperation between pigeons, SVM parameters are obtained. Finally， the optimal model is used to detect the network intrusion detection. PIO-SVM is tested with KDDcup99 network intrusion data by using Matlab. The experimental results show that the detection accuracy of PIO-SVM is better than GA-SVM and PSO-SVM, which is a practical method for network intrusion detection.

Keywords—network security; intrusion detection; Pigeon-inspired Optimization ; Support Vector Machine; parameter optimization

I. INTRODUCTION

With the development of computer network, the problem of network security is becoming more and more serious. Network intrusion has attracted widespread attention. As early as 80s twentieth Century, Anderson defined the intrusion as an unauthorized attempt to access information, tamper with information, make the system unreliable, or unable to use [1]. Therefore, the purpose of intrusion detection is to identify the malicious intent and behavior in the network. The main method is to capture and clean the data packets flowing through the network, and then use the appropriate data analysis algorithm to determine whether the packets belong to normal or abnormal data [2]. The algorithm used to determine whether the data is normal is the main point of intrusion detection. The previous intrusion detection technology is based on a pattern matching method, which is simple and easy to use, but the flexibility and adaptability are poor, and cannot recognize new intrusion behaviors. Many machine learning based on approached have been put forward, such as hidden Markov model [3] and neural network [4], which have obtained good results in the field of network intrusion detection. However, the accuracy of these methods has a great relationship with the number of samples obtained, and the fluctuation is relatively large.

Support vector machine (SVM) is a machine learning method based on structural risk minimization proposed by Vapnik et al. [5], which can be applied to intrusion detection under the condition of insufficient prior knowledge. According to the features of network intrusion data such as small samples, nonlinearity, and high dimensions, the performance of SVM is good. Mukkamala S. applied the standard SVM to intrusion detection in 2002 [6], and got better experimental results compared with the neural network. AM Chandrasekhar proposed a new approach by utilizing data mining techniques such as KNN and SVM for helping intrusion detection system to attain higher detection rate [7]. WH chen also introduces SVM and ANN into intrusion detection [8].

However, the selection of SVM parameters, including the penalty factor C, the nuclear function type and the nuclear function parameters, has a great influence on classification accuracy [9]. Studies have shown that there is a multi-peak function relationship between the performance of SVM and the parameters of the penalty factor and the kernel function [10]. At present, the commonly used SVM parameter optimization methods include genetic algorithm [11], gradient descent algorithm [12] and standard particle swarm optimization algorithm [13], however, the above optimization algorithm will fall into the local optimal solution in the optimization process, which cannot achieve the optimal classification.

Inspired by the homing behavior of pigeons in nature, Duan proposed a new swarm intelligence optimization algorithm, pigeon-inspired optimization algorithm(PIO), and applied it to autonomous flight control of unmanned aerial vehicle(UAV) cluster [14].

Sun adopted the PIO to optimize the parameters of linear quadratic controller [15], Dou utilized PIO to obtain the optimal parameters of the carrier controller [16], Deng turned the parameter design problem into parameter optimization problem and employed PIO to overcome the difficult of manually adjusting parameters in the automatic landing system [17]. In this paper, PIO is employed to optimize SVM parameters and applied to network intrusion detection.

The rest of the work is organized as follows: Section 2 gives overview of SVM. Section 3 briefly introduces the PIO algorithm. Section 4 expounds the details of the proposed approach. In section 5, the experimental results are displayed and analyzed. Finally, the work is concluded in section 6.

II. OVERVIEW OF THE SUPPORT VECTOR MACHINE

The SVM is based on structural risk minimization theory that builds the optimal hyperplane in the feature space, which allows the learning machine to be optimized, and the expectation of the whole sample space is to be satisfied by some probability.

For training samples (x_i,y_i), $i=1,...,l$ where $x_i \in R^n$ and $y \in \{1,-1\}$, SVM needs to find the optimal solution that satisfies (1):

$$\min_{\omega,\alpha,\xi} \frac{1}{2}\omega^T\omega + C\sum_{i}^{l}\xi_i \tag{1}$$

Subject to

$$y_i(\omega^T z_i + b) \geq 1-\xi_i, \xi_i \geq 0, i = 1,...l, C > 0.$$

By using the kernel function $z_i = \phi(x_i)$, the training instance x_i can be projected to higher dimensions. C is the penalty parameter of the error term; it is the upper bound of all variables. The dual problem can be obtained by using the Lagrangian multiplier method for (1).

$$\min_{\alpha} F(\alpha) = \frac{1}{2}a^T Qa - e^T a \tag{2}$$

Subject to

$$0 \leq \alpha_i \leq C, i = 1,...,l \qquad y^T a = 0,$$

Where e is the vector of all ones and Q is an l by l positive semi-definite matrix $Q_{ij} = y_i y_j K(x_i, x_j)$,

Where $K(x_i, x_j) = \phi(x_i)^T \phi(x_i)$ is called the kernel function.

The decision function is (3)

$$\mathrm{sgn}(\omega^T\phi(x)+b) = \mathrm{sgn}(\sum_{i=1}^{l} a_i y_i K(x_i, x) + b) \tag{3}$$

where

$$\omega = \sum_{i=1}^{l} a_i y_i \phi(x_i) \tag{4}$$

Gauss Radial Basis (RBF) is the mostly used kernel as (5):

$$K(\tilde{x}, \bar{x}) = \exp\left(\frac{-\|\tilde{x}-\bar{x}\|}{2\sigma^2}\right) \tag{5}$$

There are many other kernel functions, such as Linear Kernel, Polynomial Kernel, Sigmoid Kernel.

The test error expectation for a training set is therefore:

$$R(a) = \int \frac{1}{2} |\, y - f(x,a) \,| \, dP(x,y) \tag{6}$$

When a density $P(x,y)$ exits, $dP(x,y)$ may be written $p(x,y)dxdy$. This is the theoretical average error expect. But $P(x,y)$ is unknown.

$$R_{emp}(a) = \frac{1}{2l}\sum_{i=1}^{l} |\, y_i - f(x_i, a) \,| \tag{7}$$

The quantity $R(a)$ is what we called the expected risk.

Since the parameters used by the SVM, mainly depending on the instance of the support vector, are not dependent on the number of input features, the SVM does not need to avoid over-fitting this way by reducing feature input.

III. THE PIGEON-INSPIRED OPTIMIZATION ALGORITHM

PIO is a swarm intelligence optimization algorithm, which is first applied to the aerospace field through the unique navigation mode of pigeons. In order to simulate the homing behavior of dove, two operators, Map and compass operator and Landmark operator, are introduced.

（1）*Map and compass operator*: Pigeon use the height of the sun and earth field as a compass to adjust their direction. When they fly to their destinations, they have less dependence on the sun and magnetic particles.

In the optimization process, each feasible solution is used as an individual i, each individual has its position and its corresponding speed, which is recorded as $X_i=[X_{i1},X_{i2},...,X_{iD}]$ and $V_i=[V_{i1},V_{i2},...,V_{iD}]$, where $i=1,2,...,Np$. Np represents the number of possible solutions, and D represents the dimension of the feasible solution. Individual position and velocity are updated every time in D dimensional space with iteration. In the t iteration, the velocity and position of each individual follow formula as (8) (9):

$$V_i(t) = V_i(t-1)\cdot e^{-Rt} + rand\cdot(X_g - X_i(t-1)) \tag{8}$$

$$X_i(t) = X_i(t-1) + V_i(t) \tag{9}$$

where R is the map and compass factor, *rand* is a random number, and X_g is the current global best position, and which can be obtained by comparing all the positions among all the individual.

From (8) and (9), it can be seen that each individual is approaching the optimal solution while maintaining the original speed and direction. On the other hand, the position of the optimal individual is referred to, and the final direction is the vector sum of two directions. This prevents premature convergence and keeps the population diverse.

（2）Landmark operator： When the pigeons get close to their destination, they rely mainly on landmarks. When they are far away from their destination, they rely on pigeons that are familiar with the landmarks. In the landmark operator, the number of individuals is halved in each iteration, and the first half of the optimal fitness is selected as the population, and the central position of remaining individual is calculated as the X_c, which is replaced by the map and compass operator as the reference direction, and the position of each pigeon updated according to (10) (11) (12):

$$Np(t) = \frac{Np(t-1)}{2} \tag{10}$$

$$Xc(t) = \frac{\sum Xi(t) * fitness(Xi(t))}{Np(t) * \sum fitness((t))} \tag{11}$$

$$Xi(t) = Xi(t-1) + rand * (Xc(t) - Xi(t-1)) \tag{12}$$

where *fitness()* is the quality of each individual. At this point, the reference direction of the individual is the center of some superior individuals, and the population can quickly converge to the optimal solution because it does not interfere with the velocity inertia of the individual itself. When the number of iterations reaches the maximum number of cycles, the landmark operator stops operation.

IV. THE PROPOSED APPROACH

A. The idea of SVM parameters optimization

As mentioned in the previous section, SVM has two parameters including the penalty factor C and the RBF nuclear function parameters σ , so in the optimization algorithm, the dimension of each individual is two. In general, the main approach is to maximize fitness (classification accuracy) by taking two parameters.

In this paper, real number coding is adopted, each position vector of pigeon represents the candidate parameters of SVM. X_i means i-th pigeon position in the population. The fitness represented by each pigeon will be the standard for evaluating excellent individuals.

B. The Objective function for parameter optimization

The goal of SVM learning parameters is to use PIO to explore the limited subset of possible values to obtain the parameters that maximize the correct classification rate. Therefore, the support vector machine parameters optimization of objective function is the accuracy of the training data set, which can be defined as (13).

$$fitness(i) = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{13}$$

Where T_P is the number of positive considered as positive in test sets, T_N is the number of negative considered as negative in test sets, F_P is the number of positive considered as negative in test sets, F_N is the number of negative considered as positive in test sets [18].

C. Procedures of PIO-SVM

The main flow of PIO optimized SVM parameters proposed for intrusion detection is as follows:

Algorithm: PIO-SVM

1.Initialization

Initialize N_{c1max}, Nc_{2max}, Np, D, R and *search range*, which are displayed in the Table 3.

Initialize position X and velocity V_i for each pigeon individual randomly. Set $X_{pi}=X_i$, $Nc=1$

The parameters generated by PIO were passed into the SVM, and the classification accuracy was calculated and returned to PIO as the fitness value.

$X_g=argmax[f(X_{pi})]$

2.Map and compass operations

For $N_c=1$ to N_{c1max} **do**

For $i=1$ to Np **do**

While X_i is beyond the *search space* **do**

Calculate V_i and X_i according to (8) and (9)

End while

End for

Evaluate X_i and update X_{pi} and X_g

End for

3. **Landmark operations**

For $N_c=N_{c1max}$ to N_{c2max} **do**

While X_p is beyond the *search space* **do**

Sort by their fitness

$Np=Np/2$

The pigeons are discarded to half, depending on the degree of fitness sorting

$X_c=$ the center of the remaining pigeons.

Calculate X_i according to (12)

End while

End for

4.Output:

The optimal parameters and accuracy of SVM

V. EXPERIMENT RESULT AND DISCUSSION

To verify the performance of proposed PIO-SVM, three public UCI data sets and the most commonly used data set for intrusion detection (KDDcup99) are employed. Table 1 lists the details of the three UCI datasets, section A briefly introduces the basic information of the KDD data set. In addition, to demonstrate the advantage of this method, GA to optimize the SVM parameters (GA-SVM) on these data sets was also carried out in this paper. Table 4 shows the results of experiments, compared with GA-SVM and PSO-SVM, PIO-SVM has faster convergence speed and better performance for accuracy.

The proposed method is implemented in Matlab 2016b on a personal computer with a 2.30 GHz CPU, 4.00GB RAM using the Windows 7 operating system.

A. Data for experiment

To verify the method presented in this paper, three UCI public data sets are tested. The general information of 3 data sets is given in Table 1.

TABLE I. BASIC INFORMATION OF 3 DATA SETS

data sets	*attribute*	*class*	*instance*
Adult	14	2	3576
Image segmentation	19	7	2310
Letter recognition	16	7	4000

And, the most popular KDDcup99 dataset was introduced to verify the application of proposed method in intrusion detection.

The KDDcup99 data set was carry out by Defense Advanced Research Projects Agency(DARPA) at that Lincoln laboratory at the MIT Lincoln for an intrusion detection project, which mainly includes four types of intrusion. The data set has 43 attributes. Each piece of data has a label that describes the type of attack.

B. *Data preprocessing*

We divide the original data sets into five classes, one of which is normal and the other four are intrusions. To avoid bias caused by the sample imbalance, we use resampling [19]. And to divide data sets into training sets and test sets, where the training sets accounted for 70%, the test sets accounted for 30%, bootstrap sampling method is utilized [20]. Min-max normalization [21] are also adopted to improve the performance of our model.

C. *Experiment results*

SVM parameters are passed to the PIO algorithm, and the classification accuracy is used as a cost function. The parameters used is shown in Table1 (for GA) and Table 2 (for PIO). And, to make a fair comparison of the optimization capabilities of each algorithm, their initial population and algorithm iterations are the same.

TABLE II. PARAMETER USED IN GA

parameter	*explanation*	*value*
N	Number of chromosome	20
Pc	Cross ratio	0.8
Pm	Mutation ratio	0.08

TABLE III. PARAMETER USED IN PIO

parameter	*explanation*	*value*
N	Number of pigeon	20
R	map and compass factor	0.3
N_{c1}	maximum number of iterations of *map and compass operator*	18
N_{c2}	maximum number of iterations of *landmark operator*	7

TABLE IV. EXPERIMENT RESULT FOR ALL DATA SETS

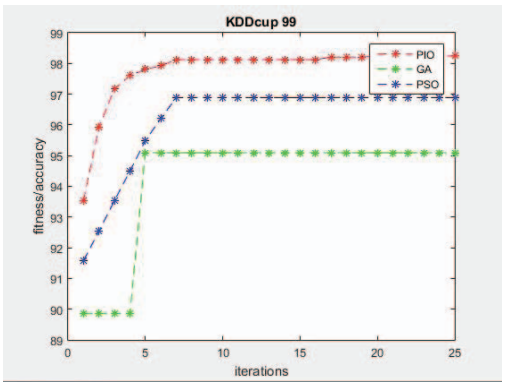
Dataset	Algorithm	Acc(%)	C	σ	Time(s)
KDD	SVM	88.43	Default	Default	
	PIO-SVM	98.24	12.6	2.3	2915
	GA-SVM	95.09	470.0	15.1	3801
	PSO-SVM	96.88	736.0	0.9	3014
Adult	SVM	80.71	Default	Default	
	PIO-SVM	84.62	16.9	0.4	3725
	GA-SVM	81.34	44.3	7.74	4950
	PSO-SVM	82.90	770.7	0.9	3814
Letter	SVM	84.80	Default	Default	
	PIO-SVM	96.26	46.4	9.0	1343
	GA-SVM	94.44	590.0	4.96	1420
	PSO-SVM	95.20	10.01	0.4	2175
Image	SVM	74.91	Default	Default	
	PIO-SVM	84.49	35.2	1.0	2406
	GA-SVM	80.26	254.0	0.03	2874
	PSO-SVM	83.36	4.01	0.01	2455

From the Table 4, it can be seen that the optimized SVM has achieved good results on each data set. Especially in the KDDcup99 data set, PIO demonstrates its superior ability to find optimum, the classification accuracy searched by PIO has reached 98.24%. And, compared with GA and PSO, PIO prevails in terms of precision and runtime.

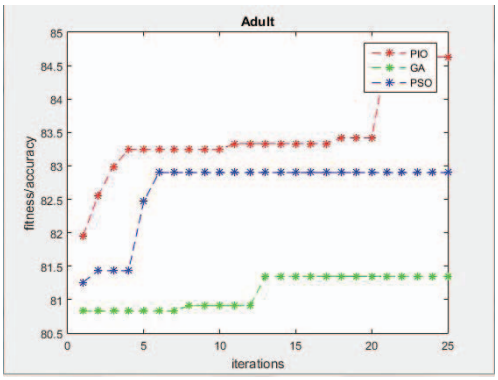
To make the experiment results more intuitive, the average fitness curve for each dataset is shown in the following figure 1. Figure 1.(a) shows the experiments on KDDcup99. Figure 1.(b) shows the experiments on Adult, Figure 1.(c) shows the experiments on Image segmentation, and Figure 1.(d) shows the experiments on Letter. It is obviously that PIO-SVM has better optimization ability compared with GA-SVM and PSO-SVM, its accuracy for SVM has the best result in all data sets.

In Adult, although GA converges faster than PIO, it is apparent that GA falls into local optimum. In Image segmentation and Letter, PIO not only shows excellent performance in accuracy, but also performs well in terms of convergence speed. The accuracy has reached 84.6%, 84.4%, respectively.

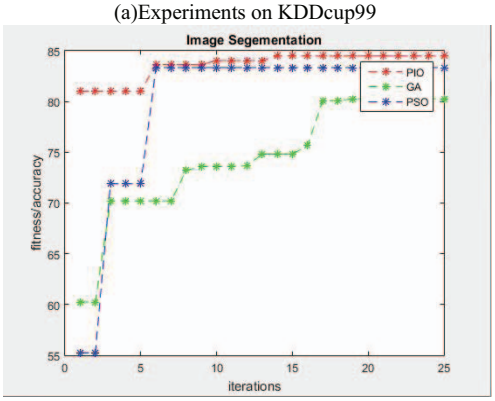
Finally, in KDDcup99, unlike GA-SVM and PSO-SVM, PIO-SVM shows excellent performance and its accuracy reaches 98.24%, which proves the feasibility of proposed method.



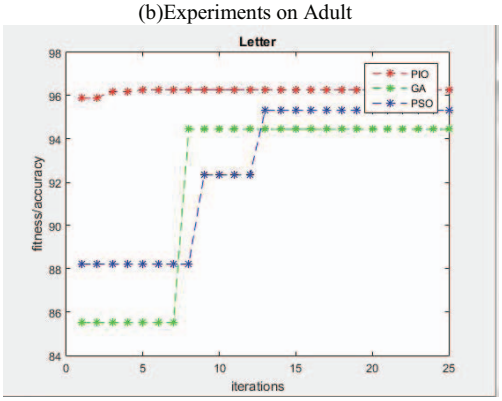
(a)Experiments on KDDcup99



(b)Experiments on Adult



(c)Experiments on Image segmentation



(d)Experiments on Letter

Figure 1. The average fitness curve for each data set

VI. CONCLUSION

In order to obtain higher classification accuracy of SVM in intrusion detection, PIO is introduced to optimize SVM parameters in this paper. Meanwhile, GA-SVM and PSO-SVM are also implemented and compared. Firstly, to verify the performance of proposed method, three UCI public data sets are tested and then intrusion detection experiments are performed on KDDcup99. The experimental results show that PIO-SVM is superior to GA-SVM and PSO-SVM in all data sets, especially in KDDcup99. Its classification accuracy reaches more than 98% in intrusion detection. In the future, the proposed PIO-SVM model will be tested on more data sets.

ACKNOWLEDGMENT

This paper is funded by the National Natural Science Foundation of China (41301371), supported by Natural Science Foundation of China (No.61772180, No.61502155)

REFERENCES

[1] J. P. Anderson, "Computer security threat monitoring and surveillance," in *James P Anderson Co., Fort*, 1980.

[2] R. G. Bace, "Intrusion detection," *Computerworld,* vol. 3, no. 1, pp. 1--29, 2000.

[3] F. Gao, J. Sun, and Z. Wei, "The prediction role of hidden Markov model in intrusion detection," in *Electrical and Computer Engineering, 2003. IEEE CCECE 2003. Canadian Conference on*, 2003, pp. 893-896 vol.2.

[4] J. Li, C. N. Manikopoulos, J. Jorgenson, and J. Ucles, "HIDE: a Hierarchical Network Intrusion Detection System Using Statistical Preprocessing and Neural Network Classification," *Proc IEEE Workshop on Information Assurance & Security,* pp. 85--90, 2001.

[5] V. N. Vapnik, "The Nature of Statistical Learning Theory," *IEEE Transactions on Neural Networks,* vol. 38, no. 4, pp. 409-409, 1996.

[6] J. Krause, M. Scalf, and L. M. Smith, "Identifying important features for intrusion detection using support vector machines and neural networks," in *Applications and the Internet, 2003. Proceedings. 2003 Symposium on*, 2003, pp. 209-216.

[7] A. M. Chandrasekhar and K. Raghuveer, *Intrusion detection technique by using k-means, fuzzy neural network and SVM classifiers.* 2013, pp. 1-7.

[8] W. H. Chen, S. H. Hsu, and H. P. Shen, "Application of SVM and ANN for intrusion detection," *Computers & Operations Research,* vol. 32, no. 10, pp. 2617-2634, 2005.

[9] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Networks the Official Journal of the International Neural Network Society,* vol. 17, no. 1, pp. 113-126, 2004.

[10] A. Unler, A. Murat, and R. B. Chinnam, "mr 2 PSO : A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification," *Information Sciences,* vol. 181, no. 20, pp. 4625-4641, 2011.

[11] E. Avci, "Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using

hybrid genetic algorithm–support vector machines: HGASVM," *Expert Systems with Applications,* vol. 36, no. 2, pp. 1391-1402, 2009.

[13] F. Melgani and Y. Bazi, "Classification of Electrocardiogram Signals With Support Vector Machines and Particle Swarm Optimization," *IEEE Transactions on Information Technology in Biomedicine,* vol. 12, no. 5, pp. 667-677, 2008.

[14] H. Duan and P. Qiao, "Pigeon-inspired optimization: a new swarm intelligence optimizer for air robot path planning," *International Journal of Intelligent Computing & Cybernetics,* vol. 7, no. 1, pp. 24-37, 2014.

[15] Y. Sun, N. Xian, and H. Duan, "Linear-quadratic regulator controller design for quadrotor based on pigeon-inspired optimization," *Aircraft Engineering & Aerospace Technology,* vol. 88, no. 6, pp. 761-770, 2016.

[16] R. Dou and H. Duan, "Pigeon inspired optimization approach to model prediction control for unmanned air vehicles," *Aircraft Engineering & Aerospace Technology An International Journal,* vol. 88, no. 1, pp. 108-116, 2016.

[17] Y. Deng and H. Duan, "Control parameter design for automatic carrier landing system via pigeon-inspired optimization," *Nonlinear Dynamics,* vol. 85, no. 1, pp. 1-10, 2016.

[18] M. Wang, Y. Wan, Z. Ye, and X. Lai, "Remote sensing image classification based on the optimal support vector machine and modified binary coded ant colony optimization algorithm," *Information Sciences,* vol. 402, pp. 50-68, 2017.

[19] A. Estabrooks, T. Jo, and N. Japkowicz, "A Multiple Resampling Method for Learning from Imbalanced Data Sets," *Computational Intelligence,* vol. 20, no. 1, pp. 18-36, 2010.

[20] B. Efron and R. Tibshirani, "[Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy]: Comment," *Statistical Science,* vol. 1, no. 1, pp. 54-75, 1986.

[21] W. Li and Z. Liu, "A method of SVM with Normalization in Intrusion Detection," *Procedia Environmental Sciences,* vol. 11, pp. 256-262, 2011.

[12] Z. A. Zhu, W. Chen, G. Wang, and C. Zhu, "P-packSVM: Parallel Primal grAdient desCent Kernel SVM," in *IEEE International Conference on Data Mining*, 2009, pp. 677-686.