An Approach to Improve the Classification Performance of Least Square Support Vector Machine

Sheetal Bairwa
Department of CS $ IT
Government Engineering College, Ajmer, Rajasthan, India

Jyoti Gajrani
Department of CS $ IT
Government Engineering College, Ajmer, Rajasthan, India

Abstract: Data mining and machine learning concepts play a vital role in dealing with large number of datasets. Various algorithms embedded within these concepts provide a wide array of mechanisms to classify the data into its specific classes. Least square support vector machine (LS-SVM) is a machine learning approach used to classify the data. Although it works better in comparison to other algorithms, the various changes sustained on each run of the program execution in its parametric values makes it quite resilient enough towards its performance.
Hence, this paper focuses on improvement of the performance level of traditional LS-SVM. In order to do so, traditional LS-SVM is integrated together with simplified particle swarm optimization technique (PSO). PSO, a swarm intelligence based, optimization approach is exploited to optimize the kernel parametric values of LS-SVM so that it not only improves its classification performance but significantly also improve the time execution of the program being run. The proposed method is exploited on the phishing dataset and experimental results are evaluated against the results of other algorithms.

Keyword: Phishing dataset, Machine learning, Least Square Support Vector Machine, Simple Particle Swarm Optimization, Schaffer function

I. INTRODUCTION

Phishing is a malicious online fraud that tricks the user by sending them spoofed emails or making them visit a website that is not legitimate but pretends to be one. The benign user when clicks on the link or visits the website, he or she is required to fill in the form that is specifically designed to trap the user. The form includes providing information related to personal details such as credit card details, bank account information, etc.

As soon, the user provides the information, it gets stored in the database of the phisher (a phisher is one who perform phishing) from where the necessary information can be grabbed and then this information is used by the phisher making others to believe that he or she is a legitimate identity.

According to [1], the term 'phishing' came into existence around 1995 when the scammers were making use of the emails luring the users to 'fish' for the information related to their finances and passwords.

On the global economy, the effect of phishing has been quite significant. According to RSA estimate, the worldwide damage due to the phishing attacks have taken a huge toll costing more than $1.5 billion in 2012 and is expected to grow more. [2]

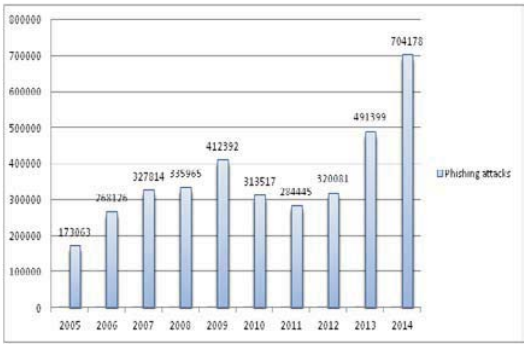

Figure 1. Statistical survey of Phishing attack

Figure 1 presents the statistical survey of phishing attacks. Accordingly, the phishing attacks continue to arise high despite of many preventive schemes. New techniques are being implemented to make phishing more sumptuous. As a result of this, phishing has been increasing at a mass scale.

This paper is sectioned as follows. Section II deals with the previous work accomplished in this field. Section III describes the least square support vector machine approach. In section IV, simple particle swarm optimization technique is explained. Section V outlays the proposed approach. Section VI exhibits the experimental results of the proposed method and last, but not the least, section VII summarizes the conclusion and future possibilities.

II. RELATED WORK

Many researchers have taken a wide interest to use phishing as their researching field. Since then, new techniques and methodologies have been adopted to provide a definite solution to the problem of phishing.

JavaScript is widely used to provide functionality such as forms submission, opening windows, carrying out input validity checks etc. At the very same time, it also provides an opportunity to the attacker to a wide range of possibilities. The problem was considered in [3]. The solution for this problem is to disengage the JavaScript but it is not viable in nature. Hence, [3] proposed a solution in the form of Anti-phish. Anti-phish was designed with the aim of deactivating the JavaScript every time the focus is set on HTML text element and reactivating it once again if focus is lost ever.

To have a win-win situation against phishing, it is essential that its detection is carried out steadfast. With this goal in mind, [4] provides a solution in the form of CANTINA. CANTINA is a content based approach used for the detection of phishing websites. The approach relies on TF-IDF information retrieval algorithm. Phisher-men make use of user identity and authentication to carry out phishing. Based on this problem, [5] proposes a solution which is based on user behavior.

Use of machine learning to detect phishing attacks was proposed in [6]. Different machine learning methods and a clustering technique was used on the dataset of phishing emails. The dataset consisted of 16 relevant features of phishing email such as HTML tags, age of domain name, IP based URL, number of domains and sub-domains, presence of JavaScript and the rest. The approach was used for the classification of phishing emails by amalgamating the basic structural characteristics in phishing emails and applying diverse algorithms. In [7], AROW was proposed. It carries out the task of extracting lexical characteristics. Afterwards, it thoroughly assesses the classification accuracy using both lexical features versus full features with a number of algorithms. The suggested modus operandi helps to enhance the accuracy and is more accurate while inflicting less computations and memory overhead.

In [8], an innovative feature extraction methodology was recommended for phishing emails that made use of latent semantic analysis and keyword extraction method. It also used K-means clustering with cosine similarity distance between the documents.
Thus, new ways are continuously being implemented either to detect the phishing activities or data mining to classify them from a pool of given dataset.

III. LEAST SQUARE SUPPORT VECTOR MACHINE

Machine Learning is a subpart of computer science evolved from computational learning theory and pattern recognition in artificial intelligence. The purpose of machine learning is to ascertain the study mechanisms and edifice of various algorithms which can learn and make different predictions on data. [9]

LS-SVM is reformulation to the standard support vector machines leading to solve the linear KKT systems. It is closely linked with the regularization networks and Gaussian techniques laying emphasis and exploiting primal-dual interpretations. Association between versions of kernel of classical pattern recognition algorithms and extension to unsupervised learning techniques, control and recurrent networks are present. Sparseness, robustness and weightings can be inflicted to LS-SVMs wherever required. For much large scale problems and on-line learning, a technique of Fixed Size LS-SVM is proposed on the basis of Nystr¨om approximation with active assortment of support vectors and approximations in the primal space.

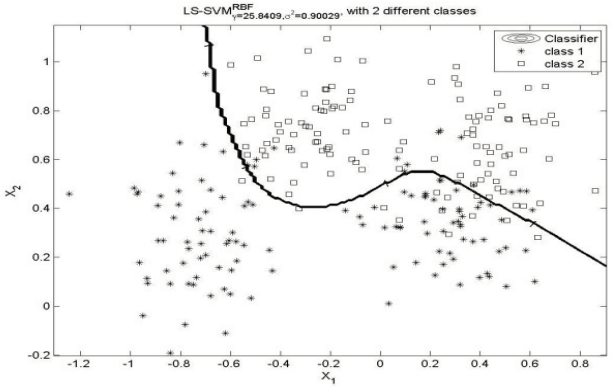

Figure 2. LS-SVM hyper-plane separating two classes

Figure 2 shows the hyper-plane of least square support vector machine that discriminates between the two specific classes using Radial Basis kernel or RBF function with parametric values gamma (γ) as 25.8409 and sigma (σ^2) as 0.90029.

A. LS-SVM Algorithm-

For a given dataset of *n* points $(x_i,y_i)^n$, i=1 with input as $x_i \varepsilon R^n$ and output as $y_i \varepsilon R$, following optimization equation is considered in primal weight space.

$$min\ J(w,b)_{w,b,e}=\frac{1}{2}w^t w+\frac{1}{2}\gamma \sum_{i=1} n\ e^2_i \qquad (1)$$

such that

$$y_i-(w^t \Psi x_i+b)=e_i, i=1,2,....n \qquad (2)$$

where γ is a factor of regularization, e_i is the difference between actual output and desired output y_i, and Ψ (.)is a nonlinear function which maps the data points into a high dimensional Hilbert space. Further, the dot product in the high-dimensional space is correspondent to a positive distinct kernel function

$$K(x_i,x_j)= \Psi(x_i)^t \Psi x_j \qquad (3)$$

In primal weight space, a linear classifier in the new space takes the following form Where w is the weight vector and bεR which is called as bias term

$$y(x)=sign(w.\Psi(x)+b \qquad (4)$$

The complete algorithm of Least Square SVM is described below.

- Load the dataset of n data points $x_i,y_i)^n$, i=1 with input as $x_i \varepsilon R^n$ and output as $y_i \varepsilon R$ consisting of target values{+1,-1}.
- Generate random weight for each data point.
- Use random values to initialize γ and σ.

- Classify the data using RBF kernel function given as

$$K(x,x_i)=\exp\left(\frac{-|x-xi|2}{\sigma^2}\right) \qquad (5)$$

- Loop the steps above until stopping criterion is reached.

One can find the solution in LS-SVM by solving a set of linear equations instead of a convex quadratic programming problem for classical SVMs. The main objective of LS-SVM is to discover an optima hyper-plane that distinguishes various classes using maximum Euclidean distance to the closest point. LS-SVM classifier maps the input vectors into a high dimensional feature space for non-separable data. Then, the LS-SVM classifier locates a hyper-plane separating the classes in this high dimensional feature space. [10]

IV. SIMPLE PARTICLE SWARM OPTIMIZATION

Particle Swarm Optimization, generally called PSO, is a swarm intelligence meta-heuristic technique to solve optimization and search space problems. Swarm intelligence is an innovative area of computer science that provides proficient computational tactics to find a solution to the problem in a manner that is quite impacted by the behaviour of real swarms. Besides, PSO, other examples are microbial intelligence, bird flocking, ant colony, fish schooling etc. [11]

PSO is a comprehensive optimization algorithm for the problems where a best solution can be depicted as a point in an n-dimensional space. The problem is optimized repetitively until a best solution is obtained.

Initially, the problem consists of a population of data. The optimization is carried out by using randomly created particles. These particles, further, are allowed to wander around the search space in accordance to a simple mathematical technique over the particle's velocity and position. The behaviour of individual particle is determined by its best known local position continuously striving unless the best available position is found in the search space.

A. Algorithm

- For every individual particle, i=1:S, initialize the position and velocity in random distributed vector manner taking in consideration the lower and upper boundaries of the search space.
- Initialize the best position of particle to its initial position as $p_i \leftarrow x_i$.
- Until the termination condition is reached, for every individual, i=1:S,

i. Evaluate the weight of each particle, ρ, given by the equation:

$$\rho(i)= \rho_{max}-((\rho_{max}-\rho_{min})/maxit)+i, \qquad (6)$$

where ρ(i)=weight of each particle,

ρ_{max} and ρ_{min} = values initialized manually, and

maxit= stopping condition

ii. For each dimension, *dim*=1:n, do,

a. Evaluate the fitness function, f, of the swarm and update the best position of swarm to $g_i \leftarrow p_i$.
b. Choose random numbers r_1 *and* r_2.
c. Update the velocity of each particle given by the equation:

$$v_{idim}= \rho(i)*v_i+r_1*(p_{idim}-x_{idim})+r2*(g_i-x_{idim}), \quad (7)$$

d. Update the position of particle by the equation:

$$x_i = x_i + v_i, \quad (8)$$

- If $f(x_i) < f(p_i)$, then,
  i. Update the best position of particle to $p_i \leftarrow x_i$
  ii. If $f(p_i) < f(g_i)$, then update the best position of swarm to $g_i \leftarrow p_i$.

- Now the optimized solution lies in the value of g.

V. PROPOSED METHOD

To further improvise the classification, this dissertation proposes a new algorithm which out beats all the above mentioned algorithms in terms of all aspects of performance. The proposed algorithm consists of two phases- Optimization Phase and Classification Phase. The optimization phase makes use of simple PSO technique and the classification phase uses LS-SVM classifier for the purpose of classifying data into different classes.
Although traditional LS-SVM can be used for classification, on being tested upon, it was found that using PSO with LS-SVM provides better results. The reason behind this is that PSO is generally effective in finding local best optimum.

The algorithm with a mix of PSO and LS-SVM is given below:

- Load the dataset X, into the working space.
- Until stopping criterion is fulfilled, do the given task:
  i. Call simple PSO_function to optimize the parameter values of LS-SVM.
- Obtain the optimized parametric values of LS-SVM classifier.
- With the optimized parametric values, train and classify the dataset into different classes.
- Evaluate the performance of the proposed algorithm in terms of true positive rate, false positive rate and accuracy and compare it with the other algorithms.

The dataset is fed into the algorithm as its input where it is first optimized using PSO algorithm. PSO algorithm, during optimization, uses a fitness function to evaluate the fitness of the swarm particles of the concerned dataset. The fitness function being used in the proposed algorithm is the Schaffer function. Schaffer function is generally a simple optimization equation generally used for optimizing the values.

Mathematically, Schaffer function is defined as: Minimize =

$$F_1(x) = \begin{cases} -x, & \text{if } x<1 \\ x-2, & \text{if } 1<x<=3 \\ 4-x, & \text{if } 3<x<=4 \\ x-4, & \text{if } x>4 \end{cases}$$

$$F2(x) = (x-5)^2$$

The optimized values so obtained are then passed to the LS-SVM classifier. Using these values, the dataset is trained and classified into distinguishable classes.

VI. EXPERIMENTATION

After the application of various algorithms including the proposed algorithm on the phishing dataset, a comparison is made to evaluate the performance of each of these algorithms. The comparison of classification performance of every algorithm for phishing dataset is tabulated below.

TABLE I. COMPARATIVE PERFORMANCE OF ALGORITHMS

Algorithm	TPR	FPR	Accuracy
Naïve Bayes	71.6	26.6	72.8
J48	89.3	10.4	89.9
Logistic Regression	68.2	32.5	67.7
Support Vector Machine	70.8	70.5	70.67
Least Square SVM	99.87	66.7	99.8
LSSVM+Simple PSO	99.96	0.0260	99.97

The ROC curve generated with this proposed algorithm is also far better than the ones generated by other algorithms.
ROCs, abbreviated for Receiver Operating Characteristics, are performance curves plotted between true positive rate and false positive rate on Y and X-axis respectively. These curves explain how the classifier performs in the regions of high sensitivity and high specificity.

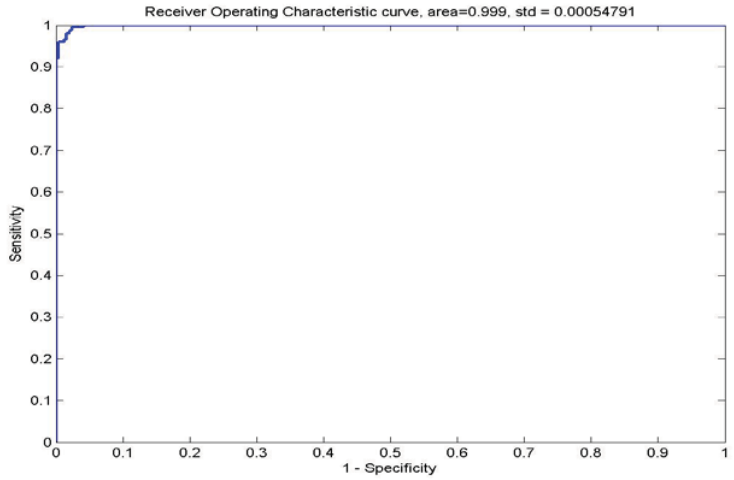

Figure 3. ROC curve for proposed approach
Figure 3 shows the ROC performance of the proposed approach.

As seen from Table II, it can be well observed that the performance of the proposed algorithm in generating ROCs is more efficient as compared to the ones generated by other algorithms.

TABLE II. ROC PERFORMANCE OF ALGORITHMS

Algorithms	ROC region of Phishing Dataset
Naïve Bayes	80.05
J48	73.6
Logistic Regression	94.7
SVM	72.3
LS-SVM	99.87
Proposed Algorithm	99.9

Given below is the bar chart of ROC curves generated with each algorithm which illustrates that the proposed algorithm is efficient in all cases.

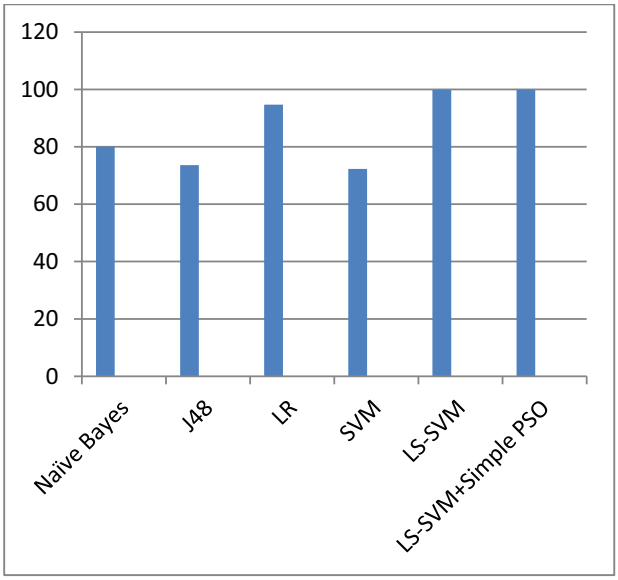

Figure 4. ROC performance for all algorithms

Further, the time execution of program is considerably minimized in contrast to that of traditional LS-SVM which significantly enhances its efficiency.

TABLE III. COMPARATIVE TIME PERFORMANCE OF LSSVM and LS-SVM + Simple PSO

Algorithm → **Dataset ↓**	**LS-SVM**	**LS-SVM + Simple PSO**
	Time Execution (in seconds)	
Blood Donation	0.22	0.19
Bank Employees	1.8	1.23
Columnar Disorder	0.8	0.1
Iris Species	0.07	0.06
Ripley	1.05	0.01
Phish URLs	0.1	0.1

VII. CONCLUSION AND FUTURE WORK

Observations from the experimental results reveal that the proposed algorithm works better in almost every aspect. The classification performance in contrast to other algorithms is also enhanced. The use of simple particle swarm optimization technique helps to achieve better results due to its ability of convergence. The particle swarm optimization technique is capable enough to find local optimum values and is the reason for its use with LS-SVM classifier.

Concluding with the paper, the future work may include improving the algorithm to make it efficient enough to classify more than two classes. Currently, the proposed algorithm distinguishes between the two classes only. Further, other optimization techniques and swarm intelligence can be used to enhance the performance.

REFERENCES

[1] Van der Merwe, A J, Loock, M, Dabrowski, "Characteristics and Responsibilities involved in a Phishing Attack", Winter International Symposium on Information and Communication Technologies, Cape Town, January 2005.

[2] Tan, ”Phishing and Spamming via IM(SPIM)”, Koontorm Center. Retrieved December 5, 2006.
[3] E. Kirda and C. Kruegel, ”Protecting Users Against Phishing Attacks with AntiPhish”,2005
[4] Y. Zhang, J. Hong, L. Cranor, ”CANTINA: A Content-Based Approach to Detecting Phishing Web sites”, ACM Proceedings, 2007
[5] Xun Dong, J.A.Clark, J.L.Jacob, “User Behaviour Based Phishing Websites Detection”, Proceedings of the International Multiconference on Computer Science and Information Technology, pp. 783-790, 2008
[6] R. Basnet, S. Mukkamala, A.H.Sung, “Detection of Phishing Attacks: A Machine Learning Approach”, Springer Proceedings, pp. 373-383, 2008
[7] A.Le, A.Markopoulou, M. Faloutsos,”PhishDef: URL Names Say It All”, arXiv publications, 2010
[8] G. Huillier, A.Hevia, R. Weber, S. Rios, “Latent Semantic Analysis and Keyword Extraction for Phishing Classification”, 2010
[9] http://en.wikipedia.org for Machine Learning
[10] X.Shao,Kun Wu, B. Liao, “Single Directional SMO algorithm for least square support vector machines”, In Proceedings of Computational Intelligence and Neuroscience, 2013
[11] http://en.wikipedia.org for PSO