A Modified Group Search Optimizer for Feature Selection and Parameter Determination of SVM

K. Joshil Raj, S Siva Sathya, Kalyan Nandi

Abstract*— Support vector machine (SVM) is a popular pattern classification method with many diverse applications. Group Search Optimizer (GSO) is a new population based optimization algorithm inspired by animal searching behavior for developing optimum searching strategies to find out solutions for continuous optimization problems. This paper presents an experimental analysis of modifications to classical GSO & studies its effects on a GSO-SVM hybrid combination for feature selection and kernel parameters optimization. In the proposed algorithm, three modifications are introduced over classical GSO to improve its global search mechanism. The quality and effectiveness of the proposed methodology has been evaluated on standard machine learning datasets.

Index Terms*— Evolutionary algorithm; Group Search Optimizer; GSO; Support Vector Machine; Machine learning; Feature Selection; Kernel parameters.

I. INTRODUCTION

Classification or supervised machine learning is a process wherein the classifier algorithm reasons from externally supplied training instances to produce a general hypothesis and makes predictions about future instances. Several algorithms have been used in the past for classification like Neural Network [1], Support Vector Machine (SVM) [2], Bayesian Networks [3], Rule based Classifiers [4], Fuzzy k-NN [5], Genetic Algorithm (GA) [6], Ant Colony Optimization (ACO) [7], and Particle Swarm Optimization (PSO) [8]. SVMs are a set of related supervised learning methods used for classification and regression. The selection of the optimal kernel parameters is of critical importance to achieve higher classification accuracies while handling learning task with SVMs. Inappropriate parameter settings lead to poor classification results [9]. Selection of optimal subsets of features from the classification data is also important for achieving higher classification accuracies Features may contain false correlations, which hinder the process of classification. Some features may be redundant since the information they add is contained in other features. . One such type of hybrid classical GSO-SVM [10] had been discussed previously by the authors.

The paper is organized as follows: Section 2 presents some of the related works. Section 3 introduces the SVM and Section 4 describes the GSO algorithm. The proposed feature selection & kernel parameter optimization of the SVM classifier with GSO is described in Section 5 and the simulation results are given in Section 6. Finally Section 7 concludes the paper.

Manuscript Received on April 2015.

K. Joshil Raj, Department of Computer Science,Pondicherry University, Puducherry, India.

Dr. S Siva Sathya, Department of Computer Science, Pondicherry University, Puducherry, India.

Kalyan Nandi, Department of Computer Science, Pondicherry University, Puducherry, India.

II. RELATED WORK

Support Vector Machines being an efficient and popular classifier, extensive amount of research has been carried out in the optimization of classification results of SVM. Zhang, Jack, and Nandi [11] in 2005 developed a GA-based approach to discover a beneficial subset of features for use of SVM for fault detection in machine condition monitoring. Pai and Hong [12] in 2006 presented a simulated annealing approach to obtain parameter values for SVM and applied it to real data. F. Alonso-Atienza et al [13] in 2012 proposed a novel feature selection algorithm based on support vector machines (SVM) classifiers and bootstrap resampling (BR) techniques. Y. Bao et al [14] in 2013 proposed an efficient memetic algorithm based on particle swarm optimization algorithm (PSO) and pattern search (PS) for SVMs parameter optimization. PSO is responsible for exploration of the search space and the detection of the potential regions with optimum solutions and pattern search (PS) is used to produce an effective exploitation on the potential regions obtained by PSO. A novel probabilistic selection strategy is also proposed to select the appropriate individuals among the current population to undergo local refinement.

For achieving true optimization in classification results it is advisable to choose algorithms that combine global and local search performing exploration & exploitation and thereby encompassing the advantages of generality, robustness, efficiency of global search and the quality of rapid convergence toward local minima of local search. Group search optimizer (GSO) is one such nature inspired algorithm
which has all these qualities and has been effectively been utilized for optimization problem solving tasks.

III. SUPPORT VECTOR MACHINE (SVM)

SVM which is an efficient and promising data classification technique proposed by Vapnik [15] in (1995), has been widely adopted in various fields of classification problems in recent years. SVM based on the tenets of statistical learning theory is now being routinely used for several binary and multi class classification tasks in different fields.

The RBF kernel is used most frequently, because it can classify multi-dimensional data, unlike a linear kernel function. RBF has fewer parameters to set than a polynomial kernel. Two parameters applied in RBF kernel of SVM are C and γ, & they need to be set appropriately. Parameter C represents the cost of the penalty. The choice of value for C influences on the classification outcome.

IV. GSO ALGORITHM

Group search optimizer (GSO) [16] is a population based optimization algorithm, inspired by animal foraging behavior. GSO Optimization employs the producer–scrounger (PS) model and the animal scanning mechanism. The population of the GSO is called a group, where each individual is called a member. A group consists of three types of members: producers, scroungers and rangers. Producers perform producing strategy in the way of animal scanning mechanism; scroungers perform scrounging strategy by joining resources uncovered by others; and rangers search for the randomly distributed resources by random walks. In each generation, the best fit member is treated as the producer, and a number of members except the producer in the group are selected as the scroungers, while the remaining members are regarded as the rangers.

Scroungers follow the producer adopting a random walk towards it. If a better position than the current producer is found by any of the scroungers then in the next searching bout it will switch to be a producer. This switching mechanism helps the group members to escape from local minima in the previous search bouts. Dispersed animals may adopt ranging behavior to explore and colonize new habitats. In each generation, rangers move to the new point based on a random head angle and a random distance.

V. THE PROPOSED GSO-SVM APPROACH FOR FEATURE SELECTION & PARAMETER OPTIMIZATION

A. *GSO Member Representation :*

Each member in GSO represents a subset of features and the parameter values of RBF kernel of SVM algorithm. Thus each GSO member will have values in '2 + n' dimensions where C & γ are the two RBF kernel parameter settings in addition to the 'n' that represents the number of features.

The individual GSO member representation is as shown below:

1	2	3				n+2
C	γ	A_1	A_2	…..	…..	A_n

n= No of Features

Where C & γ are the RBF kernel parameters settings & variable (A_1, A_2……A_n) are the features of the dataset. If the value of the variable (A_1, A_2…….A_n) is less than or equal to 0.5, then its corresponding feature is not chosen. Conversely, if the value of a variable is greater than 0.5, then its corresponding feature is chosen for making the feature subset. The searching range of parameter C is between **0.01 and 35,000** & for γ is between **0.0001 and 32.**

B. *Fitness Definition :*

Classification accuracy and the number of selected features are the two criteria used to design a fitness function. The fitness of i_{th} member is given by equation (1):

$$fitness_i = W_A \times acc_i + W_F \times \left[1 - \frac{(\sum_{j=1}^{nF} f_i)}{n_F}\right] \text{ -------- (1)}$$

Where w_A is the weight for the SVM classification accuracy. w_A is adjusted to 95%,

acc_i is the SVM classification accuracy.

w_F is the weight for the number of selected features. w_F is set to 5%.

f_i is the value of feature mask - '1' represents that feature i is selected and '0' represents that feature 'i' is not selected.

n_F is the total number of features.

C. *Three Modifications over Classical GSO-SVM:*

I. *Scrounger-Rangers Selection(SRS) Strategy 1:*

In classic al GSO-SVM the percentage of scrounger and ranger is always fixed to 80% and 20 % respectively. This proposed method 1 introduced a random change in the percentage of rangers and scrounger to strike randomness between exploration and exploitation. After every 20 iterations, percentage of scrounger is reduced by 10% and percentage of ranger is increased by 10% to introduce more randomness in the algorithm. With this procedure, the number of scroungers that will keep searching for opportunities to join the resource found by the producer will be decreased and the number of rangers will be increased thereby performing more random search strategies. This method produced results very similar to the results of classical GSO-SVM.

Algorithm : SRS Strategy 1

Initialize the population p
Initialize number of scrounger (s) = 80% of p
Initialize number of rangers (r) = 20% of p
while (i < max_iter or convergence test met){
if (i% 20==0){

$$s = s - p \times 10\%$$
$$r = p - (s + 1)$$

}
For each member, do{
Evaluate current fitness and choose the best fit individual as producer Xp of the group to perform producing
Randomly select s% members to perform scrounging
Adjust the rangers to r% to perform ranging operations
}
Set i = i+1
}

II. Scrounger-Rangers Selection(SRS) Strategy 2:

In this strategy, in the first iteration there will be one producer and 80% scroungers and 19% rangers. A counter or threshold number has been taken i.e. t1= 0.8. A random number (r1) will be generated in between 0 to 1. If r1< t1 then, increase the rangers by generating a random number r2 (whose value lies between 0 to 1). Then the number of ranger will be r2*100 of the population. The number of scroungers will be calculated by reducing the number of rangers & producer from the total population. Ranging is an initial phase of search that starts without any cues leading to specific resource thereby increasing the randomness in the behavior of the algorithm. This procedure is continued for a specified number of iterations.

Algorithm: SRS Strategy 2

Initialize the population p
Initialize number of scrounger (s) = 80% of p
Initialize number of rangers (r) = 20% of p
Threshold t1= 0.8
Generate random no r1 in between 0 to 1
while (i < max_iter or convergence test met){
if(t1> r1){
generate another random no r2 in between 0 to 1
$$s = p \times (100 - (r2 \times 100))/100$$
$$r = p - (s+1)$$
}else{
$$s = 80\%\ of\ \ p$$
$$r = p - (s+1)$$
}
For each member, do{
Evaluate current fitness and choose the best fit
Individual as producer Xp of the group to perform producing
Randomly select s% members to perform scrounging
Adjust the rangers to r% to perform rangings operations
}
Set i = i+1
}

III. Scrounger-Rangers Selection(SRS) Strategy 3:

In this strategy, a threshold value (t) is taken for the fitness value of the population & it is used to vary the number of scroungers & rangers in the population. After every iteration, all GSO member having fitness value less than t are removed and new members are generated to replace them. This method increased the probability of explorations because low fit individuals are replaced with new individuals in each iteration.

Algorithm: SRS Strategy 3

Initialize the population p
Initialize number of scrounger (s) = 80% of p
Initialize number of rangers (r) = 20% of p
Generate random no r1 in between 0 to 1
while (i < max_iter or convergence test met){

For each member, do{
Evaluate current fitness and choose the best fit individual as producer Xp of the group to perform producing
Threshold t= 60 % of fitness value of Xp
if (fitness_value _of _current _member<t){
Remove that member and introduce a new member to replace the removed GSO member
}
Randomly select s% members to perform scrounging
Adjust the rangers to r% to perform ranging operations
}
Set i = i+1 }

VI. SIMULATION RESULTS

The proposed methodology is implemented in Java using the open source java packages of WEKA 3.7.9 machine learning tool developed by Waikato University [17]. Performance is evaluated on the UCI benchmark data sets. The datasets are available at the University of California at Irvine (UCI) Machine Learning repository database [18].

A. *Experimental Settings*

The initial population of GSO is generated uniformly at random in the search space. The initial head angle Φ^0 of each individual is set to be (π/4 . . . π/4). The constant 'a' is chosen as round ($\sqrt{m+1}$) where m is the dimension of the search space i.e. number of distinct classes in training dataset. The maximum pursuit angle Θ_{max} is π/a^2. The maximum turning angle α_{max} is set to be $\Theta_{max}/2$. The maximum pursuit distance l_{max} is calculated using equation (2) given below:

$$l_{max} = \| U - L \| = \sum_{i=1}^{m} (U_i - L_i)^2 \qquad \text{---- (2)}$$

In the proposed method U_i=1 and L_i=0 for i= {1, 2…... m}.

B. *B. Experimental Result Analysis*

To measure the performance of the three different strategies the experiments were carried out on 20 standard UCI datasets. The result of the UCI datasets used in the experiments obtained without Feature Selection and with Feature selection is given in Table 1 and Table 2 respectively. Classification accuracy (%) rates obtained without Feature Selection and with Feature Selection for three different Scrounger-Rangers selection strategies have been shown in the Figure 1 and Figure 2.

TABLE I
Classification accuracy (%) rates obtained without Feature Selection for three different Scrounger-Rangers selection strategies

Data Set	GSO-SVM	SRS Strategy 1	SRS Strategy 2	SRS Strategy 3
Australian	86.52	86.52	86.52	87.68
Breast Cancer	96.84	95.13	96.99	96.99
Corral	100	100	100	100
Cleve	82.43	82.43	85.47	85.47
CRX	85.65	85.65	85.65	86.68
Diabetes	79.03	68.53	68.42	72.48
Flare	84.24	82.24	82.24	82.61
Glass	78.03	78.03	93.39	69.15
Heart	85.55	81.68	81.23	85.55
Hepatitis	89.03	85.72	82.58	85.22
German	75.04	75.50	75.20	76.20
Ionosphere	95.44	95.44	95.30	96.15
Iris	97.33	97.33	98.66	98.66
Lymphography	87.83	87.83	83.16	81.67
Pima	79.81	79.16	79.42	79.42
Soybean	93.85	87.75	93.85	95.75
Sonar	88.94	88.94	88.94	88.94
Vehicle	81.44	81.44	81.79	82.96
Vote	94.94	94.94	94.94	95.40
Vowel	99.69	99.69	99.78	99.69

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

TABLE II
Classification accuracy (%) rates obtained with Feature Selection for three different Scrounger-Rangers selection strategies

Data Set	GSO-SVM	SRS Strategy 1	SRS Strategy 2	SRS Strategy 3
Australian	87.24	87.39	87.68	90.57
Breast Cancer	97.13	96.56	97.99	97.99
Corral	100	100	100	100
Cleve	84.12	84.12	86.48	87.43
CRX	87.62	87.24	87.68	87.82
Diabetes	78.64	68.61	72.78	78.64
Flare	85.55	84.24	83.95	84.24
Glass	77.57	75.70	96.55	77.75
Heart	86.29	81.97	82.65	86.62
Hepatitis	89.03	87.09	83.24	89.03
German	75.68	76.40	77.68	77.68
Ionosphere	96.58	96.58	95.44	97.38
Iris	97.33	98.66	98.66	98.66
Lymphography	85.13	87.83	83.56	83.75
Pima	80.82	79.55	79.81	81.22
Soybean	95.75	89.89	95.83	96.42
Sonar	91.82	90.38	91.82	95.78
Vehicle	83.38	81.83	82.74	84.38
Vote	96.55	96.55	97.24	96.78
Vowel	99.79	99.69	99.79	99.79

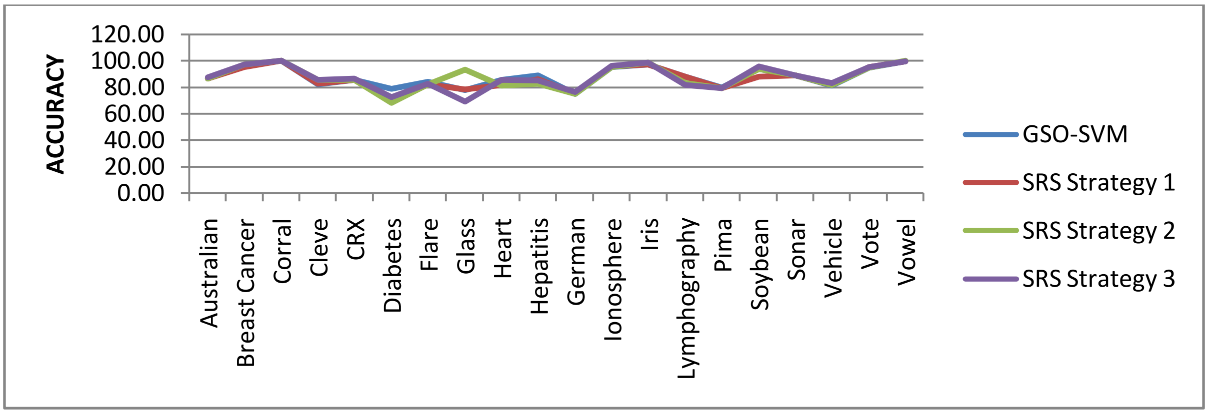

Fig.1. Classification accuracy (%) rates obtained without Feature Selection for three different Scrounger-Rangers selection strategies

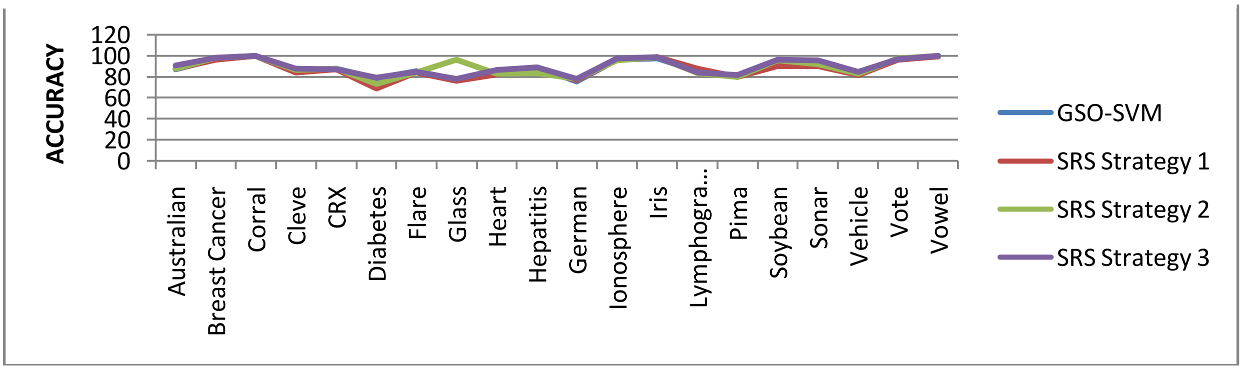

Fig.2. Classification accuracy (%) rates obtained with Feature Selection for three different Scrounger-Rangers selection strategies

VII. CONCLUSION

Classification accuracy is highly dependent upon feature selection and parameter optimization. In this paper, a modified GSO based parameter optimization & feature subset selection for a SVM classifier has been proposed. Three different scrounger-ranger selection strategies are discussed in order to improve the performance of the previous traditional GSO-SVM approach. By selecting the optimum feature subsets the dataset has been reduced in addition to selecting the optimum values of RBF kernel parameters. The UCI benchmark dataset has been used for testing purposes and remarkable increase in classification accuracy has been shown by the methodology This study has been carries out by focusing on the RBF kernel parameters, however similar procedure can be employed for optimizing other kernel parameters.

REFERENCES

1. Pang-Ning Tan, Michael Steinbach and Vipin Kumar, Introduction to Data Mining. Pearson Education Inc., 2006..
2. Han, J., and Kamber, M., Data Mining Concepts and Techniques, Elsevier Inc., Second Edition, 2006.
3. Alexandra M. Carvalho, Teemu Roos, Arlindo L. Oliveira, "Discriminative Learning of Bayesian Networks via Factorized Conditional Log-Likelihood," Journal of Machine Learning Research 12 , pp. 2181-2210,2011.
4. H. Witten, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann Series in Data Management Systems, 2005.
5. Keller, J. M. , Gray, M. R. , Givens JR. , J. A. , A Fuzzy K-Nearest Neighbor Algorithm, IEEE Transactions on Systems, Man, and Cybernetics, v. SMC-15, n. 4, pp. 258-260, 1985.
6. Kenneth A. De Jong, Evolutionary Computation A Unified Approach. The MIT Press, Cambridge, England, 2006
7. Marco Dorigo and Thomas Stützle, Ant Colony Optimization, The MIT Press Cambridge, England, 2004.
8. J. Kennedy and R. Eberhart, "Particle swarm optimization," Proc. IEEE Int. Conf. Neural Networks, pp. 1942–1948,1995.
9. Keerthi, S. S., & Lin, C.-J., Asymptotic behaviors of support vector machines with gaussian kernel. Neural Computation, 15, pp.1667–1689, 2003.
10. K. Joshil Raj, S Siva Sathya, Kalyan Nandi, GSO-SVM for Feature Selection and kernel Parameter Optimization, International Journal of Advanced Information Science and Technology(IJAIST), **ISSN: 2319 – 2682, pp. 33-39, 2014.**
11. L. Zhang, L.B. Jack, A.K. Nandi, Fault detection using genetic programming, Elsevier, Mechanical Systems and Signal Processing. 19 , pp. 271–289,2005.
12. P.-F. Pai, W.-C. Hong, Software reliability forecasting by support vector machines with simulated annealing algorithms, Elsevier Inc, The Journal of Systems and Software 79 pp.747–755, 2006.
13. Felipe Alonso-Atienza, José Luis Rojo-Álvarez, Alfredo Rosado-Muñoz, Juan J. Vinagre, Arcadi García-Alberola & Gustavo Camps-Valls, Feature selection using support vector machines and bootstrap methods for ventricular fibrillation detection, Elsevier Ltd, Expert Systems with Applications 39,pp. 1956–1967, 2012.
14. Yukun Bao, ZhongyiHu & TaoXiong, A PSO and pattern search based memetic algorithm for SVMs parameters optimization, Elsevier B.V, Neurocomputing 117, pp. 98–106, 2013.
15. Vapnik, V.N, “The Nature of Statistical Learning Theory”, Springer, Berlin Heidelberg New York ,1995.
16. S. He, Q. H. Wu, and J. R. Saunders, “Group Search Optimizer: An Optimization Algorithm Inspired by Animal Searching Behavior,” IEEE Transactions On Evolutionary Computation, vol. 13,no. 5, pp. 973-990, 2009.
17. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer,,Peter Reutemann, Ian H. Witten; The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1,2009.
18. Lichman, M., UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science,2013.

AUTHOR PROFILLE

K JOSHIL RAJ, M.Tech., received the B. Tech. degree in Electrical Engineering from the College of Military Engineering, Pune, India, in 2004 & M. Tech. degree in Computer Science & Engineering from Defence Institute of Advance Technology, Pune, India, in 2011. He is currently pursuing Ph.D in Computer Science & Engineering from Pondicherry University, Pondicherry, India. His research interests include Computational Intelligence, Bio Inspired Algorithms, Cyber Security, Neural Networks, fuzzy logic, Remote sensing & GIS.

Dr. S. Siva Sathya, PhD., received the **B. Tech.** degree in Computer Science & Engineering from Pondicherry University, Pondicherry, India, in 1993 & **M. Tech.** degree in Computer Science & Engineering from Pondicherry University, Pondicherry, India, in 1999. She is currently an Associate Professor in the Department of Computer Science, School Engineering and Technology in Pondicherry University, Pondicherry, India. Her research interests include Genetic Algorithms, Bioinformatics, Intrusion Detection & Response, Computer Based Music Analysis..

Kalyan Nandi, M.Tech., received the **B.Sc.** degree in Computer Applications (Major) from the Ramakrishna Mission Vidyamandira College, Belur, Calcutta University, West Bengal, India, in 2011 & **M.Sc**. degree in Computer Science from Pondicherry University, Pondicherry, India. He is currently doing **M.Tech.** in Computer Science and engineering from Pondicherry University, Pondicherry, India. His research interests include Bio Inspired Algorithm, Machine learning, Distributed System & Cloud Computing