The International Journal Of Engineering And Science (IJES)
||Volume||2 ||Issue|| 6 ||Pages|| 92-101||2013||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Multi-Objective Genetic Algorithm Approach to SVM Kernel Functions Parameters Selection

Kavita Aneja Saroj
Dept. of Computer Science G.J.U, Hisar, Haryana

ABSTRACT
Support Vector Machines (SVMs) have been accepted as a promising technique to address the classification problems. This method is able to achieve high predictive accuracy subject to providing appropriate values for the parameters of kernel functions. The process of tuning the kernel parameters is also known as model selection. These parameters vary from one dataset to another, therefore, SVM model selection is an important dimension of research. This paper presents a Multi-Objective Genetic Algorithm approach to optimize the kernel parameters. In this work, a MOGA is designed with two conflicting objectives to be optimized simultaneously. These two objectives are based on the error rate and a ratio of number of support vectors to the number of instances of the dataset under consideration. To evaluate the performance of the proposed method, experiments were conducted on the datasets from LibSVM (library for SVM) tool webpage and the results obtained were compared with the traditional grid algorithm for parameters searching. Compared with grid algorithm, the proposed MOGA based approach leads to less error rate and gives a Pareto front of solutions giving a user an opportunity to exercise choices with a trade-off between the two objectives.

KEYWORDS - *genetic algorithm, multi-objective genetic algorithm, parameter selection, support vector machine, kernel function.*

Date of Submission: 08 June 2013, Date of Publication: 25 June 2013

I. INTRODUCTION

A Support Vector Machine (SVM) is a statistical classification method introduced by Vapnik and coworkers in 1992 [1]. This technique has shown promising results while addressing classification problems in various fields like medical diagnosis [2], text and image recognition [3][4] and bioinformatics. SVMs are supervised learners that classify data presented with class labels by determining a set of support vectors that outline a maximal margin hyper plane. To apply SVM to non-linearly separable classification problems, data is mapped into high dimensional space using some kernel functions. These kernel functions are important as they influence the performance of the SVM. Every kernel function has parameters associated with it which govern the training process. It is important to tune these parameters including the penalty parameter C before we can really employ SVM as an effective classifier. This process of parameter tuning is known as model selection or parameter selection which aims at finding a classifier that minimizes the error in prediction [5].

The grid algorithm has been used for finding the best C and the specified kernel parameters. However, this method is time consuming and does not perform well as described by Hsu and Lin (2002). Promising results have been obtained by using Genetic Algorithms (GAs) to optimize the SVM parameters [6]. Nonetheless, a GA has the limitation of considering one objective to be either maximized or minimized. It has been observed that no single criterion is sufficient and there are number of objectives, most often conflicting in nature, that have to be taken into account for designing an effective SVM classifier.

In this work, we propose a Multi-objective Genetic Algorithm (MOGA) to find best parameter C and the other associated kernel functions parameters. Multi-objective Genetic Algorithm has the potential to simultaneously consider number of conflicting criteria and gives non- dominated solutions. The set of non-dominated solutions are known as Pareto-optimal or non-inferior solutions. In view of the fact that none of the solutions in the non- dominated set is absolutely better than the other, any one of them is an acceptable solution [7]. In this paper, two objective functions have been considered for finding the appropriate C and associated kernel function parameters. These objective functions involve error rate and risk of the SVM classifier in terms of number of support vectors and instances in the training data.

The remainder of the paper is organized as follows. Section 2 divulges into the essential details of SVM. The related work on SVM parameter selection is given in Section 3. The MOGA based approach proposed in this paper is described in Section 4. The experimental design and results are presented in Section 5. Finally, conclusions and the future direction of this work are listed in Section 6.

II. SUPPORT VECTOR MACHINES

Support Vector Machines are a relatively new learning method used for binary classification. It is also been extended for multi-class classification which is still an ongoing research issue. The basic idea in SVM classifier is to find a hyperplane which separates the n-dimensional data perfectly into two classes. However, since example data is often not linearly separable, SVM's introduce the idea of a "kernel induced feature space" which casts the data into a higher dimensional space where data become separable. Therefore, we can say SVM are of two types: one is linear SVM which does not use any kernel function and the other is non-linear SVM which use kernel functions to find a non-linear and complex decision boundary.

2.1 Linear SVM

A SVM is a supervised learning algorithm for the classification of both linear and non-linear data. Given a training set of instance-label pairs (x_i, y_i) ; i = (1, 2, 3…, m) where each $x_i \in X \subseteq R^n$ is an input vector and $y_i \in \{-1,+1\}$ is the corresponding binary class label. A SVM separates the examples by the means of Maximal Margin Hyperplane (MMH). SVM finds this hyperplane using Support Vectors which are nothing but the subset of training instances and the margins which are the sides of the hyperplane at the shortest distance. Support Vectors lie on these margins [5][8]. The algorithm strives to maximize the distance between examples that are close to the hyperplane. This margin of maximum separation is related to Vapnik-Chervonenkis dimensions (VCdim) which measures the complexity of the classifier.

The equation for a separating hyperplane is written as w.x + b = 0; where w is a weight vector associated with every attribute, namely $\{w_1, w_2, \ldots\ldots., w_n\}$; and b is a scalar referred as the bias. For linearly separable case, the data points will be correctly classified by the following equations [9]:

$$\text{For } y_i = +1,\ w^T x_i + b \geq +1 \qquad (1)$$

$$\text{For } y_i = -1,\ w^T x_i + b \leq -1 \qquad (2)$$

Equation (1) and (2) can be combined into one set of inequality as given below.

$$y_i(w^T x_i + b) \geq 1 \qquad (3)$$

Examples/Instances in the training dataset which satisfy eq. (3) with equality are called support vectors. The SVM finds the MMH by solving the following optimization problem [10]:

$$\min_{w,b} \frac{1}{2}\|w\|^2$$

$$\text{subject to:} \quad y_i(w^T x_i + b) \geq 1, \quad \forall\, i \qquad (4)$$

This is a quadratic optimization problem and to solve it one needs to find the saddle point of the Lagrange function as given below [11]:

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i [y_i (w^T . x_i + b) - 1] \qquad (5)$$

where $\alpha_i \geq 0$ is the vector of Lagrange multiplier corresponding to the constraint (4) associated with every training instance. By differentiating above eq. with respect to w and b, the following equations are obtained which are also known as Karush Kuhn-Tucker (KKT) conditions [12]:

$$\frac{\partial L(w,b,\alpha)}{\partial w} = w - \sum_{i=1}^{m} \alpha_i y_i x_i = 0, \qquad (6)$$

$$\frac{\partial L(w,b,\alpha)}{\partial b} = \sum_{i=1}^{m} \alpha_i y_i = 0 \qquad (7)$$

We also get an additional KKT complementary condition as given below:

$$\alpha_i [y_i (w^T . x_i + b) - 1] = 0, i = 1,2.., m \qquad (8)$$

The above complementary condition implies that the instances corresponding to α_i > 0 are Support Vectors. From eq. (6) and (8) we get

$w = \sum_{i=1}^{m} \alpha_i y_i x_i$ and $b = y_i - w^T . x_i$. The overall decision function can be written as follows:

$$F(x) = sign(w^T x + b)$$
$$= sign(\sum_{i=1}^{m} y_i \alpha_i (x^T x_i) + b) \qquad (9)$$

Now by substituting the eq. (6) and (7) in eq. (5), we get the final dual formulation of the eq. (4) given in following eq. (10):

$$\underset{\alpha}{Max} \ \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j x_i x_j$$
$$\text{subject to:} \quad \sum_{i=1}^{m} \alpha_i y_i = 0,$$
$$\alpha_i \geq 0, \ \ i = 1,\ldots\ldots, \ m. \qquad (10)$$

2.2 Linear generalized SVM for linearly non-separable data

It is to be noted that there will be some misclassification of examples in which linear hyperplane does not exist. To get the settings for this problem the non negative slack variables $\xi_i \geq 0$ are introduced for every example, i=1, 2, 3, ., m such that the eq. (1) and (2) becomes:

$$\text{For } y_i = +1, \ w^T x_i + b \geq 1 - \xi_i \qquad (11)$$
$$\text{For } y_i = -1, \ w^T x_i + b \leq -1 + \xi_i \qquad (12)$$

In terms of these slack variables, the problem of finding the hyperplane now becomes:

$$\min_{w, b, \xi} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{m} \xi_i$$
$$\text{subject to:} \quad y_i (w^T x_i + b) \geq 1 - \xi_i, \quad \forall i$$
$$\xi_i \geq 0 \qquad \forall i \qquad (13)$$

The penalty parameter C characterizes the weight given to the classification errors [9]. The decision function remains the same as of eq. (9) for any Support Vector x_i such that 0< α_i <C.

2.3 Non linear SVM

For the non-linear case, SVM maps the data to some higher dimensional space using a mapping function Φ, also called as kernel function. The optimization problem to be solved remains the same as eq. (13). Under this mapping the solution obtained has the following form [13]:

$$F(x) = sign(w^T \Phi(x) + b$$
$$= sign(\sum_{i=1}^{m} y_i \alpha_i \Phi(x) \Phi(x_i) + b) \qquad (14)$$

The mapping is performed by a kernel function $K(x, y) = \Phi(x).\Phi(y)$ which is the dot product of two feature vectors in decision function eq. (14). Hence the decision function f(x) given by non linear SVM becomes:

$$F(x) = sign(\sum_{i=1}^{m} y_i \alpha_i K(x, x_i) + b) \qquad (15)$$

The dual form of eq. (10) is reformulated as below:

$$\underset{\alpha}{Max} \ \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i x_j)$$
$$\text{subject to:} \quad \sum_{i=1}^{m} \alpha_i y_i = 0,$$
$$0 \leq \alpha_i \leq C, \quad i = 1,\ldots\ldots, \ m. \ (16)$$

The most common kernel functions available include linear, Radial Basis Function (RBF), sigmoid, and polynomial kernels which are given in eq. (17), (18), (19), (20) respectively [14]. In order to improve classification accuracy, the parameters for these kernel functions should be appropriately set.

Linear kernel: $\quad x * y \qquad (17)$

RBF kernel: $\exp(-g*|x-y|\wedge 2)$ (18)

Sigmoid kernel: $\tanh(g*x^T.y+r)$ (19)

Polynomial kernel: $(g*x^T y+r)\wedge d$ (20)

III. RELATED WORK

A grid approach to parameter selection is based on ideas from design of experiments which starts with a coarse grid search and refines both the grid solution and search boundaries iteratively. After each of the iteration all the sample points are evaluated and the point with best performance is chosen. The results obtained by this technique demonstrate that it is robust and works effectively and efficiently on a variety of problems [15]. A method based on crossbreed genetic algorithm, which uses two fitness functions, has been proposed to choose the kernel functions and their corresponding parameters based on two criterions namely: empirical estimators and theoretical bounds for the generalization error for improving the performance of SVMs. This method has been able to avoid premature convergence and, thereby, improves predictive accuracy [16].

Another GA-based approach was able to simultaneously optimize the SVM parameters and size of the feature subset without degrading the classification accuracy. In this approach three criteria- classification accuracy, number of selected features, and the feature cost were combined to create a single objective function. The results obtained using this approach have achieved higher predictive accuracy and that is with fewer number of features as compared to grid-based approach [6]. Gene expression programming (GEP) has also been used to evolve kernel functions. This technique trains a SVM with the kernel function most suitable for the training data set rather that pre-specifying the kernel function. Here, kernel functions were derived from the various existing kernel functions and fitness of the kernel was measured by calculating accuracy through cross validation. Results obtained using GEP also attained high accuracy and generalization capabilities for a given data set [17].

A meta-strategy utilizing genetic algorithm was proposed to automate model selection for support vector machines which combined various kernel functions in data-driven manner. In this strategy, two fitness functions were designed to represent two criteria which were incorporated in two genetic algorithms. On comparing these two model selection measures, it was found that both are appropriate to choose a classifier that will generalize well to unknown data [5]. A multi-objective optimization method has also been suggested for SVM model selection using the NSGA-II algorithm. Non -dominated set of solutions were obtained using two objectives which are false acceptance (FA) and false rejection (FR) rates. Experimental results obtained from the proposed work were analyzed and it was found that solutions achieved by employing FA/FR curve dominate the solution obtained with the criterion area under ROC curve (AUC) [18].

Based on the behavior of model parameters, Zhao et al. (2008) made an area search table and after analyzing it, they suggested first a parametric distribution model followed by a genetic algorithm based on this model to improve classification performance. The results obtained indicate that classification accuracy of genetic algorithm based on parametric distribution model is better than that of grid search [19]. Scatter search meta-heuristic approach, an evolutionary method which integrated with support vector machine called as 3SVM was devised to find near optimal values of kernel parameters. To measure the performance of the proposed method, experiments were conducted on ten datasets and the results attained were very promising and gave competitive performance as compared to other available methods [20].

IV. PROPOSED MULTI-OBJECTIVE GA APPROACH TO SVM PARAMETERS SELECTION

4.1 Multi-Objective GA

Multi-Objective GA is a multi-objective optimization algorithm that can simultaneously optimize two or more conflicting objectives. Due to conflicts among objectives, no single solution is best with respect to all objectives. For multiple objective cases, there exist a set of solutions known as non-dominated solutions or Pareto optimal solutions. This dominance concept has been proposed by Vilfredo Pareto in the 19th century. A decision vector is said to dominate another decision vector $\vec{v}$ if $\vec{u}$ is not worse than for any objective functions and if is better than for at least one objective function. This is denoted as $\prec$. More formally, a solution is said to be Pareto optimal if it is non-dominated by any other solution. Therefore, a Pareto optimal solution cannot be improved with respect to any objective without worsening any other objective. For a given Pareto optimal solution set, the corresponding objective function values in the objective space are called Pareto front.

Genetic algorithms (GAs) consider one objective which is either to be minimized or maximized. Therefore, for multi-objective optimization problems, GA, which is based upon Darwin's principle of the 'survival of the fittest', is modified as Multi-objective GA (MOGA) [21]. Generally, a multi-objective GA differs from a simple GA based on their fitness assignment procedures, and diversification or elitism approaches. Operations like selection, crossover and mutation remain same as that of GA. MOGA ranks the individuals according to the dominance rule and each solution is assigned a fitness value based on its rank, and not on its actual objective function value. Fitness sharing or crowding technique is used to maintain diversity in the population. Elitism means maintaining the best solution and is implemented by either storing it internally in the population itself or externally in a secondary list [7]. There are many MOGAs developed to date [22-27] with the purpose to find Pareto optimal solutions.

4.2 Criteria for model selection

Model selection is an optimization process which requires the choice of several efficient criteria to be optimized. The accuracy of classification and risk of classifier are often used to evaluate the performance of SVM. Therefore, we consider the following two objectives to be optimized simultaneously. The first one is error rate which needs to be minimized and is given by:

$$F(1) = 100/\text{accuracy} \quad (21)$$

The risk can be estimated by VC dimension. But the VC dimension is difficult to estimate. So we have used a simple bound T for the leave-one-out error given in [1] as our second objective:

$$F(2) = N_{sv}/N_m \quad (22)$$

Where N_{sv} is the number of Support Vectors and N_m are the number of training examples.
The whole process of the MOGA approach is shown in Fig. 1 below:

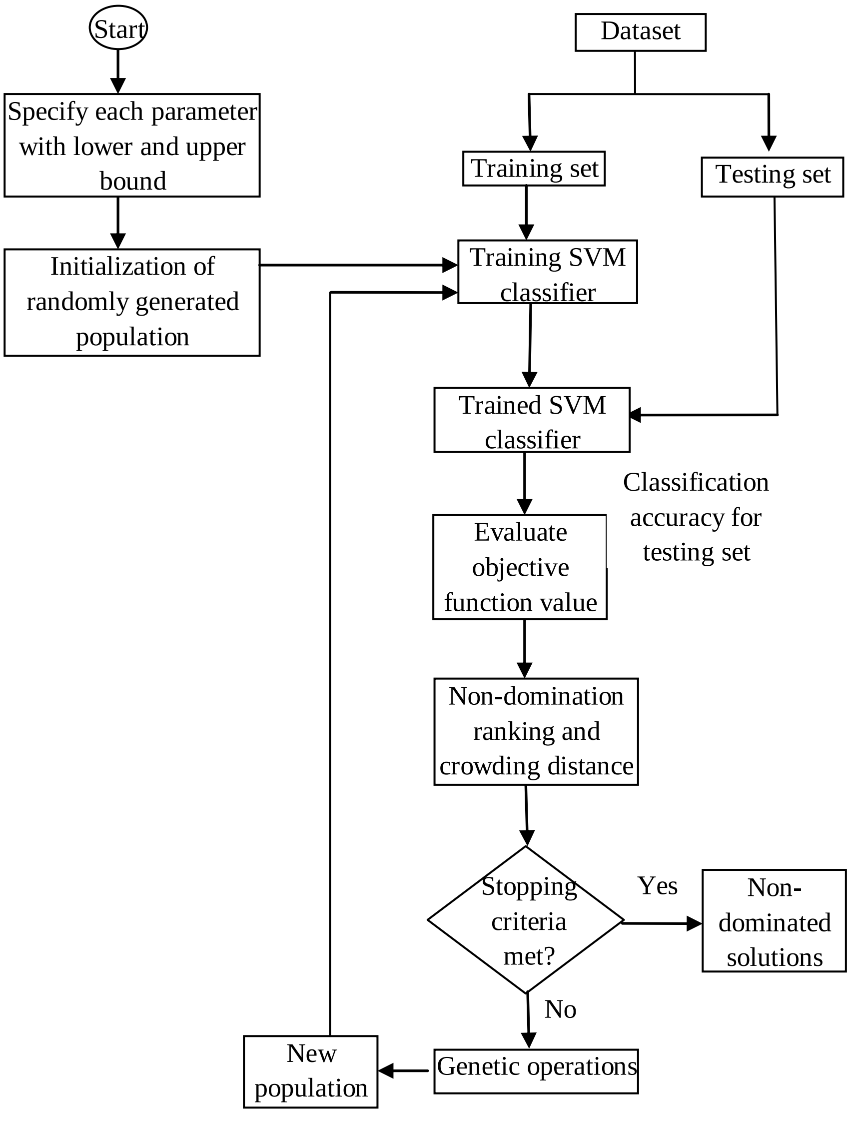

Fig. 1 The MOGA based approach

V. EXPERIMENTAL RESULTS AND DISCUSSION

We have carried out our experiments on the MATLAB 2010 using the MOGA toolbox development environment and the LibSvm. The MOGA toolbox in MATLAB consists of a variation of NSGA-II algorithm [28]. LibSvm is a library for support vector classification. A general use of LIBSVM involves two steps: first, a training data set is used to obtain a model and subsequently the model is validated on a test data set for predictive power. The empirical evaluation was performed on Intel Pentium Dual CPU running at 1.73 GHz, 3 GB of RAM and windows 7 professional Operating System. The proposed MOGA based model selection is validated on seven datasets taken from LibSvm webpage [14]. Table I describes these datasets in terms of number of attributes, instances and classes.

Table I: Datasets Information

Dataset	Features	Instance	Classes
Breast Cancer	10	683	2
Diabetes	8	768	2
Heart disease	13	270	2
Liver-disorders	6	345	2
Ionosphere	34	351	2
Australian	14	690	2
German	24	1000	2

We have considered a 2/3 partition of each dataset as training set for model construction and a 1/3 partition as the test set for evaluating the classification accuracy. For each dataset, SVM model was constructed using different a kernel functions. We have used the default settings for population size (20) and number of generations (no of parameters of kernel function * 200) of MOGA toolbox. Results of the proposed approach are compared with standard SVM classifier trained using linear, RBF and sigmoid kernel. Model selection for the standard SVM is obtained by means of grid search algorithm in which pairs of various parameters' values are tried and the one with the best cross-validation accuracy is chosen. The grid search algorithm searches the best parameters with respect to one objective only which is predictive accuracy. In fact, model selection is a multi-objective optimization problem as described in the previous sections. The ranges of parameters for the standard SVM are listed in Table II.

Table II: Parameter Range for Grid Search*

Kernel function	Log(C)	Log(g)	Log(r)
Linear	{-12,….,5}	-	-
RBF	{-12,.....,5}	{-12,….,1}	-
Sigmoid	{-12,….5}	{-12,…..1}	{0,..,2}

*All parameters are varied on log scale. A '-' sign indicates that the corresponding parameter is not present in that particular kernel.

The MOGA based SVM model selection gives non-dominated multiple solutions with respect to both the objectives and any one of them can be acceptable solution. Results obtained using linear sigmoid, radial kernels are shown in Table III, IV and V respectively. The error rate in this paper is measured as 100/accuracy which implies the minimum error rate to be one when accuracy is maximum i.e. 100. It is clear from the tables that the proposed approach has been able to achieve less error rates almost on all the datasets as compare to standard SVM with all kernel functions except on Australian dataset with linear kernel and German dataset with sigmoid kernel. It is also to be noted that if we look at results in F (1) column, it is observed that RBF performs better than other kernels and it gives more number of solutions giving more choices for user to exercise. Similarly by observing values in column F (2), we found that linear kernel performs better than the other kernels for all the datasets. The Pareto fronts obtained for various datasets using RBF kernel is shown in Fig. 2, 3, 4 and 5.

Table III: Results for Linear Kernel

Sr. no.	Datasets	MOGA		Std. SVM
		F(1)	F(2)	F(1)
1	Breast Cancer	1.0100	0.146	1.0111
2	Diabetes	1.228	0.662	1.2329
3	Heart disease	1.149	0.395	1.1739
		1.161	0.377	
4	Liver-disorders	1.468	0.787	1.4839
5	Ionosphere	1.069	0.337	1.093
6	Australian	1.195	0.341	1.174
7	German	1.29	0.54	1.3029

Table IV: Results for Sigmoid Kernel

Sr. no.	Datasets	MOGA		Std. SVM
		F(1)	F(2)	F(1)
1	Breast Cancer	1.007	0.173	1.011
		1.015	0.154	
		1.010	0.159	
2	Diabetes	1.228	1.029	1.237
		1.233	0.979	
3	Heart disease	1.161	0.673	1.213
		1.174	0.642	
		1.20	0.623	
4	Liver-disorders	1.50	1.357	1.550
		1.516	1.343	
5	Ionosphere	1.069	0.655	1.250
		1.061	0.698	
		1.053	0.703	
6	Australian	1.185	0.539	1.164
		1.19	0.529	
7	German	1.333	0.982	1.315
		1.361	0.89	
		1.356	0.94	
		1.389	0.885	

Table V: Results for RBF Kernel

Sr. No.	Datasets	MOGA		Std. SVM
		F(1)	F(2)	F(1)
1	Breast Cancer	1.007	0.149	1.014
		1	0.207	
		1.004	0.205	
2	Diabetes	1.15	0.842	1.237
		1	1.107	
		1.081	0.914	
		1.027	0.953	
		1.017	1.003	
		1.133	0.864	
		1.12	0.868	
		1.033	1.033	
		1.051	0.933	
		1	0.994	
		1.009	0.998	
		1.069	0.796	

3	Heart disease	1.059	0.827	1.1739
		1.102	0.722	
		1.029	0.877	
		1.019	0.907	
4	Liver-disorders	1.2	1.208	1.366
5	Ionosphere	1	0.679	1.0294
		1.007	0.418	
6	Australian	1.042	1.031	1.189
		1.062	0.923	
		1.078	0.758	
		1.034	1.138	
		1.15	0.539	
		1.091	0.686	
		1.066	0.836	
		1.057	0.964	
7	German	1.01	1.303	1.351
		1.003	1.47	
		1	1.488	
		1.078	1.032	
		1.05	1.127	

\

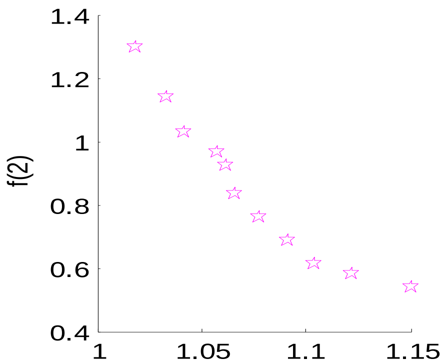

Fig. 2 Pareto front for Diabetes

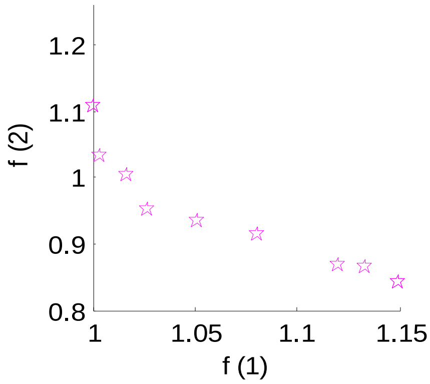

Fig. 3 Pareto front for Heart

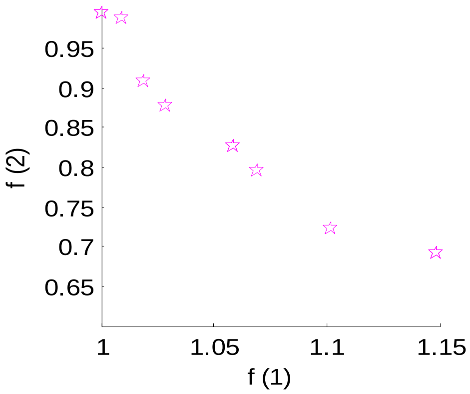

Fig. 4 Pareto front for Australian

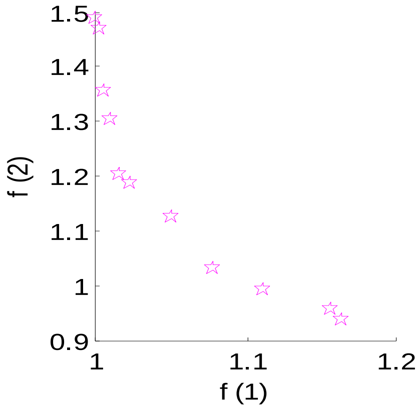

Fig. 5 Pareto front for German

VI. CONCLUSION

In this paper, we presented a MOGA based approach to SVM model selection. The kernel parameters are important to SVM and influence the performance of a SVM. But it is difficult to choose a kernel function and its parameters because they are dependent on datasets. A SVM model selection is multi-objective optimization problem, therefore, MOGA based approach has been applied to optimize the parameters of SVM according to two objectives, namely error rate and risk of classifier. We conducted experiments to evaluate the performance of the proposed approach with three different kernel functions and the results obtained were compared with those obtained with grid algorithm (standard SVM). The results obtained show enough evidence that the proposed approach has less error rates across most of the datasets as compared to grid algorithm. We can also conclude that RBF kernel gives better results as compared with other kernel functions. Further, we plan to implement the MOGA bases approach to some of the novel user-defined kernel functions [17].

REFERENCES

[1] V. N. Vapnik, *The nature of statistical learning theory* (USA: Springer-Verlag, 1995).
[2] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, Gene selection for cancer classification using support vector machines, *Machine Learning*, *46*(1-3), 2002, 389–422.
[3] T. Joachims, Text categorization with support vector machines: Learning with many relevant features, *Proceedings of the 10th European Conference on Machine Learning*, London, 1998, 137–142.
[4] G. Guo, S. Z. Li, and K. L. Chan, Support vector machines for face recognition, *Image Vision Computing, 19*(9), 2001, 631–638.
[5] S. Lessmann, R. Stahlbock, and S. F. Crone, Genetic algorithms for support vector machine model selection, *Proceedings of IEEE International Joint Conference on Neural Networks*, 2006, 3063–3069.
[6] C.-L. Huang and C.-J. Wang, A GA-based feature selection and parameters optimization for support vector machines, *Expert System Applications, 31*(2), 2006, 231–240.
[7] A. Konak, D. W. Coit, and A. E. Smith, Multi-objective optimization using genetic algorithms: A tutorial, *Reliability Engineering & System Safety, 91*(9), 2006, 992–1007.
[8] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods* (Cambridge University Press, 2000).
[9] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques, Third Edition* (Morgan Kaufmann, 2011).
[10] Prof. Olga Veksler, CS434a541a Pattern Recognition (Available at: http://www.docstoc.com/docs/44048557/CS434a541a-Pattern-Recognition-Prof-Olga-Veksler).
[11] LagrangeForSVMs.pdf. (Available at: http://www.engr.mun.ca/~baxter/Publications/LagrangeForSVMs.pdf)
[12] S. D. Sharma, *Operations Research for Hons. & Post-graduate Students* (Kedar Nath Ram Nath, 1974).

[13] S. Maldonado and R. Weber, A wrapper method for feature selection using support vector machines, *Information Sciences*, *179* (13), 2009, 2208–2217.

[14] LIBSVM -- A Library for Support Vector Machines. (Available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm/).

[15] C. Staelin, Parameter selection for support vector machines, *Hewlett-Packard Co. Tech Rep Hpl-2002-354r1*, 2003.

[16] H.-J. Liu, Y.-N. Wang, and X.-F. Lu, A method to choose kernel function and its parameters for support vector machines, *Proceedings of IEEE International Conference on Machine Learning and Cybernetics*, *7*, 2005, 4277–4280.

[17] K. Thadani, V. K. Jayaraman, and V. Sundararajan, Evolutionary selection of kernels in support vector machines, *Proceedings of IEEE International Conference on Advanced Computing and Communications*, 2006, 19–24.

[18] C. Chatelain, S. Adam, Y. Lecourtier, L. Heutte, and T. Paquet, Multi-objective optimization for SVM model selection, *Proceedings of IEEE International Conference on Document Analysis and Recognition, 1*, 2007, 427–431.

[19] M. Zhao, K. Tang, M. Zhou, F. Zhang, and L. Zeng, Model parameter selection of support vector machines, *Proceedings of IEEE Conference on Cybernetics and Intelligent Systems*, 2008, 1095–1099.

[20] M. H. Afif and A.-R. Hedar, Data classification using Support Vector Machine integrated with scatter search method, *Proceedings of IEEE Japan-Egypt Conference on Electronics, Communications and Computers*, 2012, 168–172.

[21] Z. Michalewicz, *Genetic algorithms+ data structures= evolution programs* (Springer, 1998).

[22] C. M. Fonseca and P. J. Fleming, Genetic algorithms for multi-objective optimization: Formulation, discussion and generalization, *Proceedings of the fifth International Conference on genetic algorithms*, *423*, 1992, 416–423.

[23] N. Srinivas and K. Deb, Multi-objective optimization using non-dominated sorting in genetic algorithms, *Evolutionary Computation*, *2*(3), 1994, 221–248.

[24] A. H. Dias and J. A. De Vasconcelos, Multiobjective genetic algorithms applied to solve optimization problems, *IEEE Transaction on Magnetics*, *38*(2), 2002, 1133–1136.

[25] V. Guliashki, H. Toshev, and C. Korsemov, Survey of evolutionary algorithms used in multiobjective optimization, *Problems of Engineering Cybernetics and Robotics*, Bulgarian Academy of Sciences, 2009.

[26] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II, *Lecture Notes in Computers Sciences*, *1917*, 2000, 849–858.

[27] M. S. CÎRCIU and F. LEON, Comparative study of multi-objective genetic algorithms, *Bullentin of the Polytechnic Institute of Iasi Tome Lvi, Section Automatic Control and Computer Sciences*, *1*, 2010, 35–47.

[28] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Transactions on Evolutionary Computation*, *6*(2), 2001, 182–197.