International Journal of Current Engineering and Technology
ISSN 2277 - 4106

Available at http://inpressco.com/category/ijcet

Research Article

A Genetic Algorithm Approach to Kernel Functions Parameters Selection for SVM

Kavita Aneja[a*], Saroj[a] and Jyoti[a]

[a]Dept. of Computer Science G.J.U., Hisar, Haryana

Accepted 24 June 2013, Available online27June 2013, **Vol.3, No.2 (June 2013)**

Abstract

The Support Vector Machines (SVM) is a classification algorithm with many diverse applications. The SVM has many parameters associated with it which influences the performance of the SVM classifier. In this paper, we employ Genetic Algorithm based approach to find and select an appropriate kernel function and its parameters. This proposed technique combines predictive accuracy and complexity of SVM as two criteria into a fitness function for evaluating the performance of SVM. Our method is compared with grid algorithm and the experimental results validate that the proposed approach is much better than the grid method.

Keywords: *Support Vector Machines, kernel function, linear kernel, RBF kernel, parameters selection, genetic algorithm*

1. Introduction

Support Vector Machine (SVM) are supervised learners first introduced by Vapnik in 1992 (N. E. Ayat, M. cheriet, and C.Y. Suen, 2005; B. E. Boser et al, 1992). In the following years, this technique has received considerable attention in various applications including pattern recognition, text and image recognition (I. Guyon et al, 2002; J. Zhang and Y. Liu, 2004), bioinformatics, medical diagnosis (T. Joachims, 1998; G. Guo et al, 2001), and the support of corporate decision making (S. Viaene et al, 2001).

SVM classifies data with different class labels by determining a set of support vectors that gives a hyperplane. This algorithm provides a mechanism to map the data in some higher dimensional space to make the given data linearly separable and this mapping is performed by a mapping function also called kernel function. Every kernel function has some well-defined parameters associated with it. These parameters should be properly set along with a penalty parameter C before SVM is actually used as a classifier. This process of finding the appropriate parameter values is known as parameter selection or model selection.

A number of methods have been proposed by researchers in literature for fixing/tuning the SVM parameters. The traditional method of finding the SVM parameters is Grid algorithm. However, this method is time consuming and does not perform well as defined by (C. W. Hsu and C. J. Lin, 2002; S. M. LaValle et al, 2004). The standard gradient descent approach is used by (O. Chapelle and V. Vapnik, 2002; Y. Tan and J. Wang, 2004), which may get trapped in some local optima and cannot ensure optimal solution. Another method a parametric distribution model for selection of parameters is proposed which is found to be better than grid search but it obtains best parameters of only one kernel function RBF (M. Zhao et al, 2008). A simulated annealing method presented to obtain parameter values of SVM for software reliability forecasting (P. F. Pai and W.C. Hong, 2006). There are no techniques for finding the optimal values for the SVM parameters and the problem is still a topic of more research to present new methods for tuning SVM parameters.

In this work, a GA based approach with existing two criteria by (H. J. Liu et al, 2005), has been combined based on its accuracy and complexity of the SVM for selecting the SVM parameter values associated with every kernel function. The concept of GA was developed by Holland and his colleagues. They are meta-heuristics algorithms that imitate the long-term optimization process of biological evolution for solving optimization problems (D. E. Goldberg and J. H. Holland, 1988).

The remainder of the paper is organized as follows. Section 2 describes the essential details of SVM. Our GA based approach is illustrated in section 3. The numerical results of experiments are presented and discussed in Section 4. Conclusions are given in section 5.

2. Support Vector Machine

The SVM is a supervised learning method for solving linear and non-linear classification problems. Given

*Corresponding author: Kavita

training set with d number of instances $\{x_i, y_i\}_{i=1}^{d}$ where $x_i \in R^n$ is an input vector and $y_i \in \{-1,+1\}$ its corresponding binary class, the idea of this method is to separate given data by the means of maximal marginal hyperplane. This hyperplane is defined by the set of support vectors which are the subset of training examples. SVM finds this hyperplane by minimizing the norm of the vector w under the constraint that the training examples having different class label lies on the opposite side of this hyperplane (S. Lessmann et al, 2006), as shown in following Fig. 1

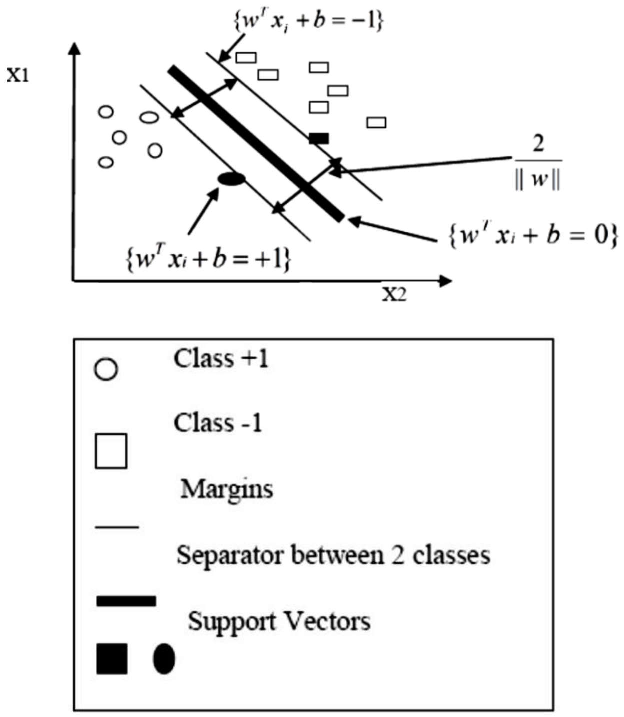

Fig. 1 Linear separation of two classes in 2-D plane

For finding the hyperplane, following optimization problem to be solved:

$$\min_{w,b} \frac{\|w\|}{2}.$$

Since $y_i \in \{-1,+1\}$, we can formulate constraint as

$$y_i((w \bullet x_i) + b) \geq 1, i = 1,2,3,\ldots,d \tag{1}$$

Examples which satisfy above constraint with equality are known as support vectors. For those instances which don't satisfy the constraint (1), SVM introduces slack variables ξ_i for misclassify examples. A key feature in SVM is that it maps the data to higher dimensions when the given data is not linearly separable by some mapping function called as kernel function $(x \rightarrow \Phi(x))$. At this time the SVM requires the solution of the following optimization problem:

$$\min_{w,b,\xi}\left(\frac{\|w\|}{2} + C\sum_{i=1}^{d}\xi_i\right)$$
$$subject\ to: y_i(w^T\Phi(x_i) + b) \geq 1 - \xi_i \tag{2}$$
$$\xi_i \geq 0$$

Where C is the penalty parameter which allows user to control the trade-off between maximizing the margin and classifying the training set without error. The optimization problem of (2) is convex quadratic optimization problem which can be solved by using Lagrange multiplier method. The primal form of the eq. (2) is as follows:

$$L(w,b,\xi_i;\alpha_i,\beta_i) = \frac{\|w\|}{2} + C\sum_{i=1}^{d}\xi_i - \sum_{i=1}^{d}\alpha_i[y_i(w^T\Phi(x_i)+b) - 1 + \xi_i] - \sum_{i=1}^{d}\beta_i\xi_i$$

Where α_i, β_i are Lagrange multipliers associated with every instance.

The coefficients α_i can be found by solving the dual form of eq. (3) which is given as follows:

$$\max_{\alpha}\left(\sum_{i=1}^{d}\alpha_i - \frac{1}{2}\sum_{i=1}^{d}\sum_{i=1}^{d}\alpha_i\alpha_j y_i y_j \Phi(x_i)^T\Phi(x_j)\right) \tag{4}$$

$$Subject \sum_{i=1}^{d}\alpha_i y_i = 0,\ 0 \leq \alpha_i \leq C$$

A mapping function or kernel function is defined as $K(x_i, x_j) = \Phi(x_i)^T\Phi(x_j)$ .So the problem is changed to

$$\max_{\alpha}\left(\sum_{i=1}^{d}\alpha_i - \frac{1}{2}\sum_{i=1}^{d}\sum_{i=1}^{d}\alpha_i\alpha_j y_i y_j K(x_i, x_j)\right) \tag{5}$$

The various types of available kernel functions include linear kernel, radial basis function (RBF) kernel and sigmoid kernel which are shown as functions in eq. (6), (7), (8) respectively (C. C. Chang and C. J. Lin, 2011). In order to improve classification accuracy, these kernel parameters in each kernel function should be properly set.

Linear kernel: x*y (6)
RBF kernel: exp (-g*|x-y|^2) (7)
Sigmoid kernel: tanh (g*x^T.y + r) (8)

3. GA based approach for searching SVM kernel functions parameters

3.1 Overview

Genetic Algorithms (GA) work with a set of candidate solutions called a population. Based on the Darwin's principle of 'survival of the fittest', the GA obtains the optimal solution after a series of iterative computations (Z. Michalewicz, 1988). A solution of GA is represented as chromosome. GA can deal with large search spaces efficiently and has fewer chances to get local optimal solution than other algorithms. Each solution is evaluated by a fitness function. The crossover and mutation functions are the main operators that affect the fitness value. The fitter chromosomes are selected for reproduction using the roulette wheel or the tournament selection methods. The whole process of GA is described in Fig. 2 below.

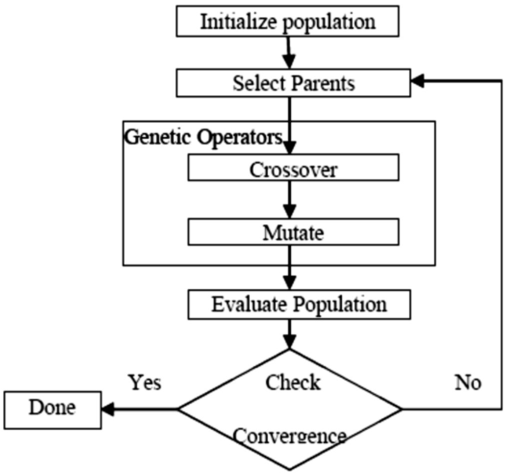

Fig. 2 Working of GA

3.2 Optimize method

Selection of SVM parameters is an optimization process, therefore GA based approach is proposed in this text. In this work, two existing criteria for evaluating the performance of SVM are combined to form fitness function which is to be minimized describe as follows:

$$F = 0.75 * \frac{1}{C_k} + 0.25 * \frac{SV}{d} \qquad (9)$$

Where C_k is the classification accuracy on the testing data set and the second criteria measures the complexity of the classifier which consist of ratio of number of support vectors (SV) and number of training examples (d).

4. Experiments Results

We implement the proposed approach using Matlab GA toolbox kit which is embedded in global optimization toolbox. The experiments were carried out on Intel Pentium Dual-Core CPU running at 2 GHz, 3GB of RAM and windows 7 professional Operating System. To study the performance of the proposed approach preprocessed datasets from LibSVM tool webpage were used (C. C. Chang and C. J. Lin, 2001). Table I lists information about these datasets.

Table I Datasets Information

S.NO.	Dataset	Attributes	Instances	Classes
1	Breast Cancer (Wisconsin)	10	683	2
2	Diabetes	8	768	2
3	Fourclass	2	862	2
4	Ionosphere	34	351	2
5	Vehicle	18	846	4

For each dataset, SVM model was constructed using two different kernel functions. In this work, parameters of both kernels (Linear and RBF) are optimized along with penalty parameter C. Results of this approach were compared with those obtained with grid algorithm. The results obtained using linear and RBF kernels are shown in Table II, III respectively. Both the tables show the optimized values of fitness function and parameters associated with kernel function. It can been clearly observed from these tables that the proposed approach gives much better results as compared with grid algorithm except vehicle with linear kernel .By looking at results of GA based approach in both the tables, it can also be concluded that linear kernel performs better than RBF kernel for first two given datasets (medical field) and RBF performs better for rest three datasets.

Table II Results for Linear Kernel

S.No	Dataset	Fitness Value (F)		Optimize parameter value
		GA	Grid	C
1	Breast Cancer (Wisconsin)	0.0148	0.7545	11.447
2	Diabetes	0.0449	1.066	10.035
3	Fourclass	1.1029	1.4112	3.647
4	Ionosphere	0.8786	0.9366	4.846
5	Vehicle	1.068	1.063	8.52

Table III Results for RBF Kernel

S.No	Dataset	Fitness Value (F)		Optimize parameter value	
		GA	Grid	C	G
1	Breast Cancer (Wisconsin)	0.0179	0.783	2.977	0.088
2	Diabetes	0.06	1.0774	25.705	0.445
3	Fourclass	0.7635	0.7986	31.12	3.85
4	Ionosphere	0.8342	0.9169	21.55	0.0727
5	Vehicle	0.9159	1.0159	757.97	0.2368

Conclusions

The kernel functions are important to SVM as they affect the performance of SVM. A Genetic Algorithm approach is proposed to choose the best kernel and its parameters. Comparison of the obtained results of the proposed approach with the grid algorithm proves that the approach suggested in this paper gives much better results. In this work, only two kernel functions’ parameters were optimized, however GA can also deal with other kernel functions’ parameters. In future, the proposed approach can be tested on more number of datasets or other real-world problems. We can also extend this work by employing Multi-objective GA (A. Konak, 2006), instead of GA.

Acknowledgements

The author would like to thank Chih-Chung and Chih-Jen Lin that Support Vector Machines were constructed with LibSVM (version 3.14) by them and for kind sharing of their skill and datasets for this problem.

References

N. E. Ayat, M. Cheriet, and C. Y. Suen (2005), Automatic model selection for the optimization of SVM kernels, *Pattern Recognition*, vol. 38, no. 10, pp. 1733–1745.

B. E. Boser, I. M. Guyon, and V. N. Vapnik (1992), A training algorithm for optimal margin classifiers, in *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, USA.

I. Guyon, J. Weston, S. Barnhill, and V. Vapnik (2002), Gene selection for cancer classification using support vector machines, *Machine Learning*, vol. 46, no. 1–3, pp. 389–422.

J. Zhang and Y. Liu (2004), Cervical cancer detection using SVM based feature screening, in Proceedings of the 7th Medical Image *Computing and Computer-Assisted Intervention*, France.

T. Joachims (1998), Text categorization with support vector machines: Learning with many relevant features, in Proceedings of the 10th European Conference on Machine Learning, Germany.

G. Guo, S. Z. Li, and K. L. Chan (2001), Support vector machines for face recognition, *Image Vision Computing*, vol. 19, no. 9, pp. 631–638.

S. Viaene, B. Baesens, T. Van Gestel, J. A. K. Suykens, D. Van den Poel, J. Vanthienen, B. De Moor, and G. Dedene (2001), Knowledge discovery in a direct marketing case using least squares support vector machines, *International Journal of Intelligent Systems*, vol. 16, no. 9, pp. 1023–1036.

C. W. Hsu and C. J. Lin (2002), A simple decomposition method for support vector machines, *Machine Learning*, vol. 46, no. 1–3, pp. 291–314.

S. M. LaValle, M. S. Branicky, and S. R. Lindemann (2004), On the relationship between classical grid search and probabilistic roadmaps, *International Journal of Robotics Research*, vol. 23, no. 7–8, pp. 673–692.

O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee (2002), Choosing multiple parameters for support vector machines, *Machine Learning*, vol. 46, no. 1–3, pp. 131–159.

Y. Tan and J. Wang (2004), A support vector machine with a hybrid kernel and minimal Vapnik-Chervonenkis dimension, *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 4, pp. 385–395.

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

P. F. Pai and W. C. Hong (2006), Software reliability forecasting by support vector machines with simulated annealing algorithms, *Journal of Systems and Software*, vol. 79, no. 6, pp. 747–755.2006

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

D. E. Goldberg and J. H. Holland (1988), Genetic algorithms and machine learning, *Machine Learning*, vol. 3, no. 2, pp. 95–99.

S. Lessmann, R. Stahlbock, and S. F. Crone (2006), Genetic algorithms for support vector machine model selection, in *Proceedings of IEEE International Joint Conference on Neural Networks*, pp. 3063–3069.

CS434a541a Pattern Recognition Prof. Olga Veksler. Available at: http:/ /www. docstoc.com/ docs/44048557/CS434a541a-Pattern-Recognition-Prof-Olga-Veksler.

C. C. Chang and C. J. Lin (2011), LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27.

Z. Michalewicz (1988), Genetic algorithms+ data structures= evolution programs. *Springer*.

C. C. Chang and C. J. Lin (2001), LIBSVM -- A library for support vector machines. Available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

A. Konak, D. W. Coit, and A. E. Smith (2006), Multi-objective optimization using genetic algorithms: A tutorial, *Reliability Engineering and System Safety*, vol. 91, no. 9, pp. 992–1007.