TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.1, January 2014, pp. 831 ~ 836
DOI: http://dx.doi.org/10.11591/telkomnika.v12i1.3182

Support Vector Machine Optimized by Improved Genetic Algorithm

Changsheng Xiang*[1], Zhou Yu[2], Xilong Qu[1]
[1]Department of Computer and Communication, Hunan Institute of Engineering, Xiangtan 411104, China, Ph./Fax:0731-58683909/ 58683586
[2]School of Information and Engineering, Henan Institute of Science and Technology, Xinxiang, 453003, China, Ph./Fax:0373-3040346/3040873
*Corresponding author, e-mail:cx5243879@sohu.com, Zhouyu@163.com, quxilong@126.com

Abstract

Parameters of support vector machines (SVM) which is optimized by standard genetic algorithm is easy to trap into the local minimum, in order to get the optimal parameters of support vector machine, this paper proposed a parameters optimization method for support vector machines based on improved genetic algorithm, the simulation experiment is carried out on 5 benchmark datasets. The simulation show that the proposed method not only can assure the classification precision, but also can reduce training time markedly compared with standard genetic algorithm.

***Keywords**: support vector machine, genetic algorithm, parameter optimization, cross-validation*

1. Introduction

Support vector machine (SVM) is a new machine learning method derived from the statistical learning theory. Since the late 1990s, SVM has shown growing popularity and has been successfully applied to a number of applications [1]. Established on the theory of structural risk minimization(SRM) principle, SVM has some distinct advantages over ANN, such as, it is globally optimal, small sample-size, good generalization ability, and it is resistant to the over fitting problem [2]. Although SVM has become more widely used in many areas, it is difficult to build a highly effective model before the parameters of SVM are carefully determined. Ni and Chapelle [3], [4] stated that the optimal parameters of SVM play a crucial role to establish a classify model with high classify accuracy and stability. To make an efficient SVM model, the parameters of SVM have to be carefully predetermined. Chen [5] has used genetic algorithm (GA) to optimize the SVM parameters, the results showed that the GA not only has reduced computation time, but also has improved the prediction accuracy. But the initial population of GA is produced randomly, which often leads to poor local search ability and it are difficult to get satisfactory results [6].

In order to solve the parameters optimization problem of SVM, this paper proposes a new parameters optimization method (IGA_SVM) which introduces weighted depth-first search strategy into initial population of GA. The results of simulation experiments on five benchmark datasets show that IGA_SVM can not only find the optimal parameters, but also reduce time complexity and *can assure the classification precision*.

2. Research Method

2.1. Principle of SVM

Let y(x_i, y_i), i=1,2…,N, where $x_i \in R^d$ is an input vector, $y_i \in \{-1,1\}$ is its corresponding desired output, n is the number of training data, and d denotes the number of dimensions of input data. In SVM, the original input space is mapped into the high dimensional space called feature space, by nonlinear mapping $x \rightarrow g(x)$, Let f(x) be the output of SVM corresponding to the input vector x, in the feature space, then a linear function is constructed:

$$f(x) = w \cdot x_i + b \tag{1}$$

Received May 10, 2013; Revised August 15, 2013; Accepted September 5, 2013

Where, w is a coefficient vector and *b* is a threshold [7].

If a hyper-plane exists that satisfies *Eq*.(1), and then linear separation is obtained. In this case, *w* and *b* can be rewritten as follows. *Eq*.(1) becomes

$$\min_{1\le i\le N} y_i(< w\cdot x_i >= b) \ge 1, i = 1,2,\cdots,N \tag{2}$$

Let the distance from the data point to the hyper-plane be 1/||w||. Among separating hyper-planes, there exists one optimal separating hyper-plane (OSH), and the distance between two support vector points on two sides of this hyper-plane is maximal. Because the distance between two support vector points is 1/||w||^2, the minimal distance to OSH, ||w||^2, may be derived from *Eq*.(2) [8].

The margin of a separating hyper-plane, calculated as ||w||^2, determines the hyper-plane's generalization ability. The OSH has the largest margin among separating hyper-planes. ||w||^2 is minimized with *Eq*.(2) and Lagrange's polynomial. Let a denote (a_1,…,a_N). Combining Lagrange's polynomial (in the order of N) with *Eq*.(2) produces the following equations for maximization.

$$W(a) = \sum_{i=1}^{N} a_i - \frac{1}{2}\sum_{i,j=1}^{N} a_i a_j y_i y_j x_i x_j \tag{3}$$

Where, $a_i \ge 0$ and under constraint $\sum_{i=1}^{N} y_i a_i = 0$.

Quadratic programming method can be adopted to solve the above maximization problem. If a vector $a^0 = (a_1^0,\cdots,a_N^0)$ satisfies the *Eq*.(3) in maximization, then the OSH expressed in terms of (w_0,b_0) may be expressed as follows:

$$w_0 = \sum_{i=1}^{N} a_i^0 y_i x_i \tag{4}$$

Where, the support vector points must comply with $a_i^0 \ge 0$ and Eq.(2).

When considering expansion in constraint *Eq*.(4), the determinant function of hyper-plane is expressed as follows:

$$f(x) = sign\left(\sum_{i=1}^{N} a_i^0 y_i x_i x + b_0\right) = 0 \tag{5}$$

In most cases, the data are not linearly separable, and are consequently mapped to a higher-dimensional feature space. Therefore, if the data cannot be classified clearly in the current dimensional space, then the SVM will map them to a higher dimensional space for classification.

Input data are mapped to a higher dimensional feature space by plotting a nonlinear curve. The OSH is constructed in the feature space. By constructing the feature space, φ(x) can be adopted in constrained *Eq*.(3) as shown below:

$$W(a) = \sum_{i=1}^{N} a_i - \frac{1}{2}\sum_{i=1}^{N} a_i a_j y_i y_j \phi(x_i)\phi(x_j) \tag{6}$$

Given a symmetric and positive kernel function *k*(x,y), the existence of Mercer's theory can be deduced. Therefore, *k*(x,y)= φ(x) φ(y). Provided that the kernel function *k*(x,y) satisfies Mercer's theory, the derived training algorithm is guaranteed for minimization.

$$W(a)=\sum_{i=1}^{N}a_i-\frac{1}{2}\sum_{i=1}^{N}a_ia_jy_iy_jk(x_i,x_j) \qquad (7)$$

The decision function is expressed as follows:

$$f(x)=sign\left(\sum_{i=1}^{N}a_iy_ik(x_i,x_j)+b\right) \qquad (8)$$

Several kernel functions help the SVM in obtaining the optimal solution. The most frequently used such kernel functions are the polynomial, sigmoid and radial basis kernel function (RBF) [9], [10]. The RBF is generally applied most frequently, because it can classify multi-dimensional data, unlike a linear kernel function. Additionally, the RBF has fewer parameters to set than a polynomial kernel. RBF and other kernel functions have similar overall performance. Consequently, RBF is an effective option for kernel function. Therefore, this study applies an RBF kernel function in the SVM to obtain optimal solution. The RBF kernel function is defined as follow:

$$K(x,y)=\exp(-\frac{\|x-y\|^2}{2\sigma^2}) \qquad (9)$$

Where, σ is the kernel width of RBF.

Two major RBF parameters applied in SVM, *C* and σ must be set appropriately. Parameter *C* represents the cost of the penalty. The choice of value for C influences on the classification outcome. If *C* is too large, then the classification accuracy rate is very high in the training phase, but very low in the testing phase. If *C* is too small, then the classification accuracy rate unsatisfactory, making the model useless. Parameter σ has a much greater influence on classification outcomes than *C*, because its value affects the partitioning outcome in the feature space. An excessively large value for parameter σ results in over-fitting, while a disproportionately small value leads to under-fitting [11], [12].

Grid search is the most common method to determine appropriate values for C and σ [13], [14]. Values for parameters C and σ that lead to the highest classification accuracy rate in this interval can be found by setting appropriate values for the upper and lower bounds and the jumping interval in the search. Nevertheless, this approach is a local search method, and vulnerable to local optima. Additionally, setting the search interval is a problem. Too large a search interval wastes computational resource, while too small a search interval might render a satisfactory outcome impossible. In addition to the commonly used, grid search, other techniques are employed in SVM to improve the possibility of a correct choice of parameter values. Particle swarm optimization algorithm (PSO) is used to optimize parameters of SVM.

2.2. Genetic Algorithm

Genetic algorithm (GA) was formally introduced in the United States in the 1970s by John Holland at Michigan University [15]. GA is often viewed as function optimizer, although the range of problems to which GA has been applied is quite broad. To use GA, you must represent a solution to your problem as a genome. And then the GA creates a population of solutions and applies genetic operators such as mutation and crossover to evolve the solutions to find the best one(s).This presentation outlines some of the basics of GA. The three most important aspects of using GA are: definition of the objective function, definition and implementation of the genetic population, definition and implementation of the genetic operators. Once these three have been defined, the generic genetic algorithm should work fairly well. Beyond that you can try many different variations to improve performance, find multiple optima species, or parallelize the algorithms.

2.3. IGA_SVM

Firstly, SVM parameters is taken as a population of GA, the population is decoded to parameters of SVM to obtain classification rates (C*R*) of test set, and then calculate corresponding GA fitness (fitness function is selected based on the values of CR directly in this paper), secondly, the next generation population is produced by selection, crossover and mutation, etc, and the progeny population is used to repeat the above operation until GA stop conditions are met, finally, the optimal parameter is obtained and the optimal model of SVM is established. The processes step of the IGA_SVM algorithm is as follows:

Step 1: initialization population. Firstly, the range of SVM parameters are determined, and then the combining weighted depth-first search is used to produce the initial population of GA with roulette selection method, the operations are as follows:

(1) A weight is assigned to every individual in the population.

(2) The depth-first search method is carried out to select a certain node by roulette selection method according to node weights.

(3) If a path which meets the requirements is found, the path is selected as a chromosome of initial population and subtracts 1 from the weights of all sides on the path.

(4) Step (2) and (3) are repeated until the numbers of chromosomes meet the size of population.

Step 2: the initial population is encoded to the parameters of SVM, the training set are input into SVM to train and built the classify model, and the built model is tested by the test dataset, and the overall test sample recognition rate is taken as the fitness of the GA.

Step 3: selection, crossover and mutation operation are carried out and produce the next generation population.

Step 4: the news parameter is obtained to train and test SVM model again and calculates fitness, if the training stop criterion of GA is meet, return to step5, otherwise return to step3.

Step 5: the parameter(c, σ) obtained at the moment is the parameter of final SVM model.

3. Results and Discussion

3.1. Data Set

LIBSVM software tool package and IGA program are used to carry out tests on IBM System x3800. The datasets are form UCI data which are a2a, diabetes, waves, w1a, heart_disease [16]. The data are divided into training set and test set, which can not only provide enough data for model learning but also verify the effect of the model. The data is as follow.

Table 1. The data set

data name	training samples	test samples
a2a	2000	1000
diabetes	500	200
waves	2500	10000
w1a	680	187
heart_disease	1000	190

3.2. Comparison Models and Evaluation Criterion

In order to make the classification results of IGA_SVM comparable, three comparison models are chosen: SVM based on grid search, SVM based on standard genetic algorithm (GA_SVM) and SVM based on particle optimization algorithm (PSO_SVM). Thee classification rate (C*R*) of test data is taken as the evaluation standard. C*R* is defined as follows:

$$CR = \frac{the\ number\ of\ correctly\ classified\ samples}{the\ total\ number\ of\ samples} \quad (10)$$

3.3. The Results Comparison of Model

The research test adopts 10-fold cross validation method, in the process of test, each dataset is divided into 10 equal subsets randomly, select one subset at random from the 10 subsets as test set and the rest subsets are conflated as training set, such tests must be done ten times and use the percentage of average recognition rate *CR* to evaluate the performance of the model. The parameters of SVM obtained for SVM, GA_SVM, PSO_SVM, IGA_SVM are shown as table 2 and the results of models are shown as figure 2.

Table 2. The parameters of models

Data set	SVM		GA_SVM		PSO_SVM		IGA_SVM	
	C	σ	C	σ	C	σ	C	σ
a2a	83.19	1.06	28.54	0.53	93.27	3.90	78.05	8.20
diabetes	6.97	0.08	50.39	0.59	67.64	1.72	233.47	4.57
waves	6.96	1.29	117.66	1.81	244.57	7.42	236.89	1.43
w1a	10.03	3.35	178.81	0.91	46.72	17.20	84.49	0.57
heart_disease	3.07	0.42	13.74	1.87	215.73	23.53	121.60	2.18

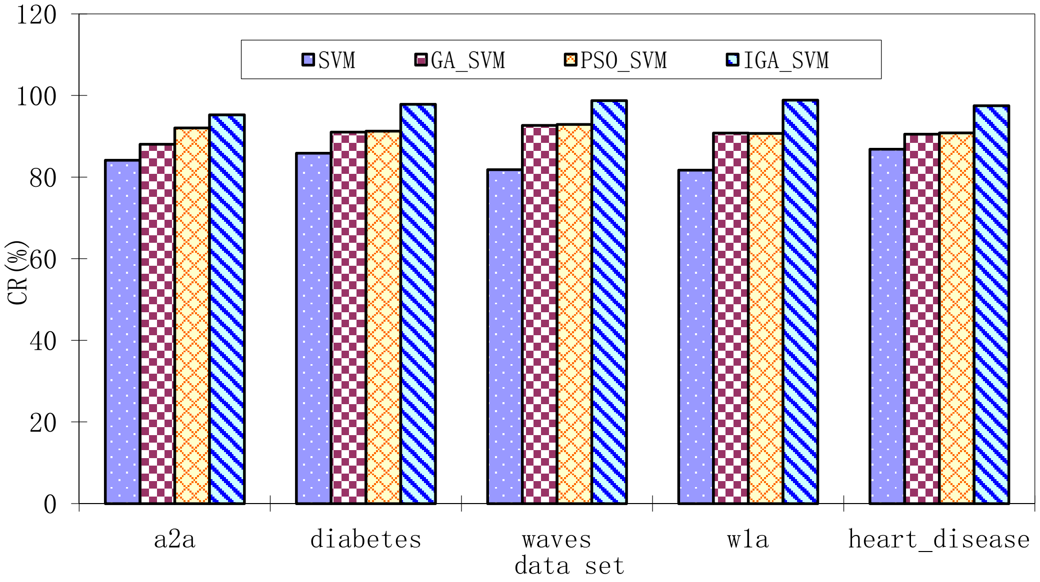

Figure 2. The CR(%) of models

It can be concluded from Table 2 that for dataset a2a, the *CR* of PSO_SVM is superior to that of SVM, the *CR* of IGA_SVM and GA_SVM are superior to SVM and PSO_SVM, IGA_SVM has the best classify results, the results shows that IGA_SVM can keep the diversity of initial population, enhance parameter optimization ability and improved *CR* due to the introduction of weighted depth-first search and roulette selection method, which shows the parameter optimization strategy of IGA_SVM is reasonable and feasible.

In Matlab 2012, the Tic and Toc is used to count training time (ms) and testing time (ms) of the all models, the results are shown in table 3. Compared the SVM, GA_SVM, PSO_SVM, the training and testing time of IGA_SVM significantly shortened, and classification efficient is improved, the experiment results show that the IGA_SVM can find the optimal the parameters of SVM, and reduces the support vectors of the SVM and computational complexity is deceased.

Table 3. The training times of different models

Data set	SVM	GA_SVM	PSO_SVM	IGA_SVM
a2a	7.51	2.57	2.18	1.28
diabetes	5.19	2.28	1.94	1.35
waves	4.86	2.36	2.09	1.71
w1a	4.28	2.55	1.84	1.81
heart_disease	5.62	1.79	1.52	1.04

4. Conclusion

Parameters plays a key role in support vector machine, if parameters optimization is unreasonable, it often causes under-learning and over-learning, thereby directly influencing prediction precision and computing time. Traditional GA has disadvantages of premature phenomenon, poor local optimization ability and slow convergence speed, etc, and this paper proposes IGA_SVM which uses weighted depth-first search and roulette selection method to ensure the diversity of initial population for genetic algorithm. The simulation results show that IGA_SVM has a good robustness and a better convergence speed, and it can improve the classify performance of SVM.

Acknowledgement

I would like to express my special thanks to referees for their precious advice to my paper. Authors gratefully acknowledge the projects supported by Provincial Natural Science Foundation of Hunan (13JJ9022) and Project supported by Provincial Science & Technology plan project of Hunan (2013GK3029).

References

[1] Sanchez AD. Advanced support vector machines and kernel methods. *Neuro Computing*. 2009; 12(5): 55: 5-11.
[2] Sapankevych NI, Sankar R. Time series prediction using support vector machines*: a survey. IEEE Computational Intelligence Magazine*. 2009; 22(4): 24-38.
[3] Ni LP, Ni ZW, Li FG. SVM model selection based on ant colony algorithm. *Computer Technology and Development*. 2007; 17(5): 95-101.
[4] Chapelle O, Vapnik V. Choosing multiple parameters for support vector machines. *Machine Learning*. 2002; 46: 131-138.
[5] Chen PW, Wang JY, Lee HM. *Model selection of SVM using GA approach*. Proceedings of the International Joint Conference on Neural Networks. 2004; 39: 36-41.
[6] Chen RC, Hsieh, Chung H. Web page classification based on a support vector machine using a weighed vote schema. *Expert Systems with Applications*. 2006; 31(12): 427-435.
[7] Han FQ, Li HM, et al. A new incremental support vector machine algorithm. *Telkomnika.* 2012; 10(6): 1171-1178.
[8] Yu Y, Zhou L. Acoustic emission signal classification based on support vector machine. *Telkomnika*. 2012; 10(5): 1027-1032.
[9] Cao LJ, Francis EH. Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on Neural Networks*. 2003; 14(10): 1506-1518.
[10] Zhang YL, Guo N, Du H, Li WH. Automated defect recognition of C-SAM image in IC packaging using support vector machines. *International Journal of Advanced Manufacturing Technology*. 2005; 25: 1191-1196.
[11] Pal M, Jayaraman VK, Kulkarni BD. Support Vector Machines for Classification in Remote Sensing. *International Journal of Remote Sensing*. 2005; 26(5): 1007-1011.
[12] Huang CL, Wang CJ. A GA-based feature selection and parameters optimization for support vector machines, *Expert Systems with Applications*. 2006; 31(12): 231-240.
[13] Wang J, Wu X, Zhang C. Support vector machines based on k-means clustering for real-time business intelligence systems. *International Journal of Business Intelligence and Data Minin*g. 2005; 21(1): 54-64.
[14] Zhang XF, Zhao Y. Application of Support Vector Machine to Reliability Analysis of Engine Systems. *Telkomnika*. 2013; 11(7): 3352-3560.
[15] Pai PF. System reliability forecasting by support vector machines with genetic algorithms. *Mathematical and Computer Modeling*. 2005; 43(11): 262-274.
[16] Yuan ZM, Zuo B, Tan XS, et al. Eexperimental design and analysis of the optimal fermentation medium based on uniform design and support vector regression*. The Chinese Journal of Process Engineering.* 2009; 12(9): 148-152.