TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol. 13, No. 1, January 2015, pp. 159 ~ 165
DOI: 10.11591/telkomnika.v13i1.6692

Nonlinear Classifier Design Research Based on SVM and Genetic Algorithm

Wenjuan Zeng*, Haibo Gao
School of Information Science and Engineering, Hunan International Economics University, Changsha, 410205, China
*Corresponding author, e-mail: matlab_bysj@126.com

Abstract

This paper presents a support vector machine (SVM) model structure, the genetic algorithm parameters of the model portfolio optimization model, and used for non-linear pattern recognition, the method is not only effective for linear problems, nonlinear problems apply effective; the law simple and easy, better than the multi-segment linear classifier design methods and BP network algorithm returns the error. Examples show the efficiency of its recognition of 100%.

Keywords: *genetic algorithm, support vector machine, pattern recognition, nonlinear, combinatorial optimization.*

1. Introduction

Support vector machines SVMs are statistical learning theory the content of the youngest and the most useful part. Its core content is presented from 1992 to 1995 [5], mainly for pattern recognition problems. SVMs is separable from the case of the linear optimal separating plane made, that such a classification hyperplane can not only correct classification of all training samples, but also from the classification of the training sample surface at the closest point to the classification of the maximum distance from surface . By making the maximum interval to control the complexity of classification, thus achieving a good generalization ability. Linear case can not be separated, through a nonlinear transformation to transform the input space into a high dimensional feature space, in this new space to strike the optimal linear classification surface. In order to avoid high-dimensional feature space in the complex non-linear operations, in SVMs kernel function method is adopted, it is the high-dimensional inner product space is converted to the original calculation of the kernel function space.

Genetic algorithm referred to as GA (Genetic Algorithms) in 1962 by the American University of Michigan Professor Holland's simulate the natural genetic mechanisms and biological evolution from a parallel random search optimization method. It "survival of the fittest, the fittest generation" of biological evolution principle into series optimized parameters of a code group, according to the value of the selected fit function and replication through inheritance, crossover and mutation screening of individuals to make adaptive value high individual was retained, to form a new group, new group inherits the previous generation of information, but also better than the previous generation. This cycle, groups of individuals to adapt to ever-increasing, until it meets certain conditions. The algorithm is simple, parallel processing, can get global optimal solution.

We used a linear pattern recognition methods and applications mature, we are probing for nonlinear problems with multi-stage or more of the linear approach; [1] is a method of research scholars, though this can often to achieve the purpose of the sample can be divided, but the method of complex structure, deal with the problem "surgery" more cumbersome.

2. Support Vector Machines for Classifier

Assume that training data $\{x_i,y_i\}$ $(i=1,\ldots,l;\ x_i\in R^n;\ y_i\in\{-1,+1\})$, Can be a hyperplane $(wx) + b = 0$ no error to separate the two samples with the largest distance separating hyperplane will

Received September 16, 2014; Revised November 14, 2014; Accepted December 12, 2014

get the best generalization ability. Optimal hyperplane will be the nearest point of a small number of samples (called support vectors) to decide, nothing to do with the other samples. Use the following form for the description and classification of the sample interval Δ hyperplane:

$$(w.x)+b=0,\ ||w||=1 \quad (1)$$

$$y=1,\ 若(w.x)+b\geq\Delta \quad (2)$$

$$y=-1,\ 若(w.x)+b\leq-\Delta \quad (3)$$

The optimization problem of SVMs in the separating hyperplane to make the following normalization: Let Δ= 1, and w and b can be scaled. Hyperplane from the nearest sample points (support vectors) satisfy:if y=1 then:

$$(w.x_i)+b=1 \quad (4)$$

$$\text{if } y=-1 \text{ then } (w.x_i)+b=-1 \quad (5)$$

Support Vector to the hyperplane distance $1/||w||$。Thus, the mathematical optimization problem formulation is:

$$\begin{aligned} &\min\ \tfrac{1}{2}||w||^2 \\ &\text{s.t.}\ \ y_i(w.x_i+b)-1\geq 0\ \ (i=1,\ldots,l) \end{aligned} \quad (6)$$

According to the most optimal solution of quadratic programming theory, the problem can be transformed into the Wolfe dual problem to solve. Lagrange function is constructed:

$$L(w,\alpha,b)=\tfrac{1}{2}||w||^2-\sum_{i=1}^{l}\alpha_i\, y_i(w.x_i+b)+\sum_{i=1}^{l}\alpha_i\ \ (\alpha_i\geq 0;\ i=1,2,\ldots,l) \quad (7)$$

Where: α_i is a Lagrange multiplier. According to optimization theory are:

$$\partial L(w,\alpha,b)/\partial w=0 \quad (8)$$

$$\partial L(w,\alpha,b)/\partial b=0 \quad (9)$$

$$w=\sum_{i=1}^{l}\alpha_i\, y_i.x_i \quad (10)$$

$$\sum_{i=1}^{l}\alpha_i\, y_i=0 \quad (11)$$

Substituting back the two Lagrange function, eliminating w and b, by computing the original optimization problem by Wolfe dual problem:

$$\begin{aligned} &\max\ \ w(\alpha)=\sum_{i=1}^{l}\alpha_i-\tfrac{1}{2}\sum_{i,j=1}^{l}\alpha_i\,\alpha_j\, y_i.\, y_j\, x_i\, x_j \\ &\text{s.t.}\ \sum_{i=1}^{l}\alpha_i\, y_i=0\ \ (\alpha_i\geq 0;\ i=1,2,\ldots,l) \end{aligned} \quad (12)$$

The original optimization problem solution is the overall optimal solution. Optimization algorithm can be solved α; Parameter b according to the Karush-Kuhn-Tucker conditions obtained:

$$b=y_i-w^T x_i\ \ (\alpha_i\in(0,C)) \quad (13)$$

Then the optimal hyperplane as:

$$f(x)=(w.x)+b=\sum_{i=1}^{l}\alpha_i\, y_i(x_i.x)+b \quad (14)$$

For linear classification problems can not be separated, you can map the input x through the nonlinear function into a high dimensional feature space $\Phi(x)$, In this space then a linear classification. The end result is, by a kernel function K (xi, x) instead of Equation (14) in $(x_i.x)$。 The optimal split plane is:$f(x)= \sum_{i=1}^{l}\alpha_i\, y_i K(x_i,x)+b$

The discriminator is:

$$f(x)= \mathrm{sgn}\{\sum_{i=1}^{l}\alpha_i\, y_i K(x_i,x)\} \quad (15)$$

Kernel function method in support vector machines played an important role, and the choice of kernel function satisfy the Mercer conditions, as long as any symmetric function.

3. The Basic Operation of the Genetic Algorithm

1) Copy (Reproduction Operator): is selected from an old population and strong vitality of the individual bits. The process of generating new populations of strings. According to the fitness value of the copied bit string, which is the mean value with a high fit in the bit string is more likely to produce the next generation of one or more children. It mimics the natural phenomenon, the application of Darwin's theory of the fittest generation. Copy operation can be achieved by random method. If using a computer program to implement, consider first produce between 0 and 1 uniformly distributed random numbers, if the probability of a string of 40% of replication, when generating random numbers between 0 and 0.4, the string is copied, Otherwise be eliminated.

2) Cross (Crossover Operator): copy operation from the old population who choose the good, but not create new chromosomes. The simulated cross-breeding process of biological evolution to imagine, through the combination of the two chromosomes exchange, to create new varieties. The process is as follows: Choose two chromosomes in the pool match, exchange of one or more points were randomly selected positions: the exchange of the right of the parent chromosomes exchange some point, you can get two new chromosome numbers string. Exchange reflects the nature of information exchange ideas. Cross a little cross, multi-point crossover, as well as the same cross, order crossover and cycle crossover. That cross is the most basic method, used widely. It refers to the chromosomes have a cut off point.

3) Variation (Mutation Operator): mutation operation used to simulate the biological environment in the natural genetic factors due to various accidental genetic mutation, it is very small probability value of random genetic change. If the only selection and crossover, and no variation, combination of genes can not be outside the initial search space, the evolution into a local solution at an early stage and into the termination process, thus affecting the solution quality. In order to obtain the greatest possible control of the quality of higher optimal solution, mutation operation must be used.

4. Model Parameters Based on Genetic Algorithm Optimization Principle

4.1. The Parameters and that

First determine the parameter range that is generally given by the user, and then by the accuracy requirements, its encoding. Select the binary string to represent each parameter, and establish the relationship between parameters. Then connect the binary string to form a long binary string, the string for the genetic algorithm can operate the object.

4.2. Select the Initial Population

Because of the need to implement the process of programming, so the initial population randomly generated by computer. For binary encoding, the first between 0 and 1 generated uniformly distributed random number, then the provisions of the random number generated between 0 and 0.5 to 1 between the representatives on behalf of 0,0.5 1. Of course, real random numbers can also be used. In addition, taking into account the complexity of computation to specify the size of population.

4.3. Determination of fitness function

General constrained optimization methods to meet the conditions can be obtained under a set of parameters in the design parameters from the group to find the best.

4.4. Genetic Operations

First fit value obtained by fitness function, and then find a copy of each string corresponds to the probability. Copy the string of each generation probability and the product of the number of strings that should be replicated in the number of the next generation. The probability of a large copy in the next generation will have more offspring, on the contrary will be eliminated.

Second, single-point crossover, crossover probability P_c. After copying a member from the P_c of the probability of selection in order to match the string form pool, pool matches and then a member of the random match, random crossover position is determined.

Finally, the mutation probability P_m. The initial population through reproduction, crossover and mutation are the new generation of population, the population on behalf of future generations into the adapter by the decoding function of the end meets the conditions of observation, if not met, then repeat the operation until satisfied.

End of the condition set by the specific issues, as long as the target parameters within prescribed limits, then terminate the calculation. The above operation can be expressed in Figure 1.

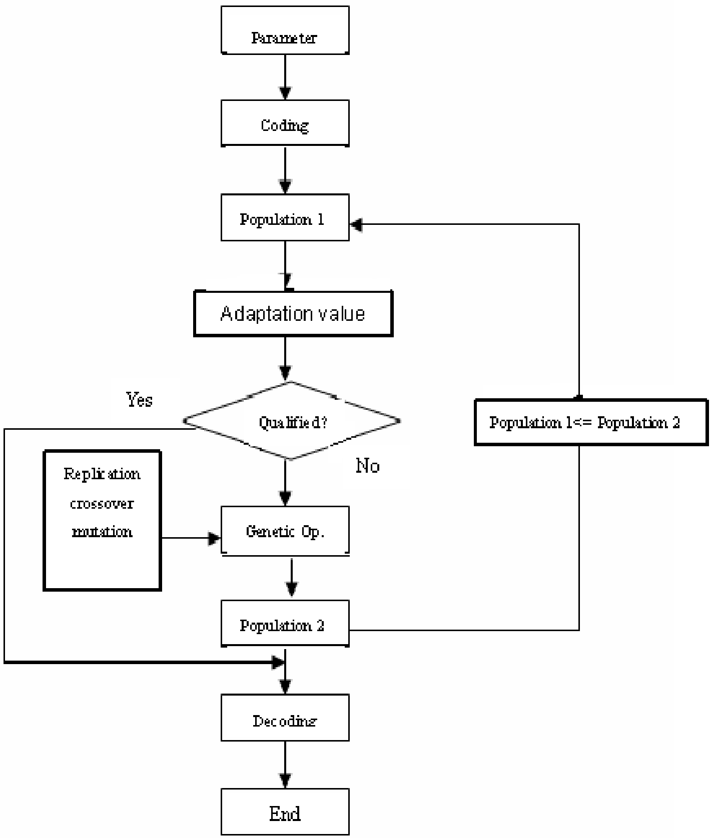

Figure 1. Flowchart of GA

4.5. Optimization of Model Parameters using Genetic Algorithm Specific Steps

1) to determine the approximate range of each parameter and the code length, coding;

2) randomly generated initial population of n individuals form P (0);

3) the population of each individual decoded into the corresponding parameter value, use this argument to the value of cost function and fitness function value of J f, taking f = 1 / J;

4) The application of reproduction, crossover and mutation operators on the population P (t) operation to produce the next generation population P (t +1);

5) Repeat steps (3) and (4). Until parameter convergence or to achieve the desired targets.

5. Example Implementation Process

5.1. Identify the Sample Data with Support Vector Machine Model

Select [1] can be divided in a nonlinear two examples, this case is two dimensional problem, it has 45 sets of samples, a class of 24 groups of samples, sample groups of 21 class 2; their raw data shown in Figure 2 (C_1 Class 1 data, that corresponds to 1 point on the graph; C_2 for the 2 categories of data, the corresponding figure on the 2; the first column of the i samples of serial order), Figure 2 [1], the multi-stage linear classification Control subdivision plans. Figure and data from the literature [1].

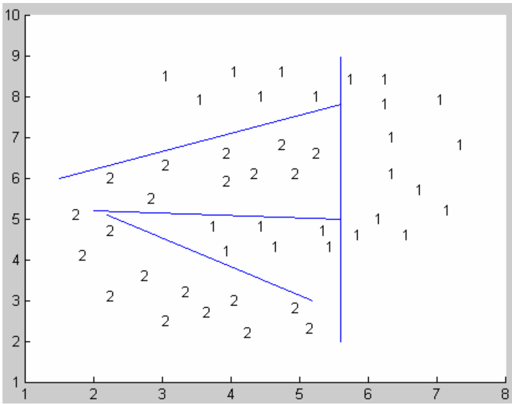

Figure 2. Original data distribution and multi-segment linear classification

As can be seen from Figure 2, the case is non-linear pattern recognition problem, we take the sensor core functions:

Sensor $\tanh(x)=(e^{x}-e^{-x})/(e^{x}+e^{-x})$
Kernel function $K(x,x_i)=\tanh(-||x- x_i||^2/2)$ (16)

In (12), using the kernel function $K(x,x_i)$ is Instead of $(x_i.x)$, And with the penalty function method, the (12) into the optimization problem:

$$\text{Min } w(\alpha)=\tfrac{1}{2}\sum_{i,j=1}^{l}\alpha_i \alpha_j y_i. y_j K(x_i, x_j)-\sum_{i=1}^{l}\alpha_i$$
$$+M^*|\sum_{i=1}^{l}\alpha_i y_i |, \quad (\alpha_i\geq 0; i=1,2,\ldots,l) \quad (17)$$

Where M is the penalty coefficient, taking M = 100. l is the sample number, the cases of l = 45. Discriminant function classification is still (15), and:

$$b= y_i -\sum_{j=1}^{l}\alpha_j y_j K(x_i, x_j) \ (\alpha_i\in(0,C)) \quad (18)$$

5.2. Parameter Optimization Genetic Algorithm

Here we use real-coded genetic algorithm, 45 samples of the cases, find the model (17), the optimal solution, where the model parameters α_i有45个。have 45. Selected for each $\alpha_i \in(0,1)$ uniform random number between the initial weight, as one species each α_l, $\alpha_i$the number of initial population of 50;_Determine the cost function value J of the rules: J= $w(\alpha)$; Crossover probability and mutation probability were P_c=0.9, P_m=0.033。With real number encoding, after 40 generations of evolution, α_i to obtain the optimal parameters in Table 1, b =0.1133.

Table 1. Optimization of model parameters α_i

1	2	3	4	5
0.3588	0.4310	0.2240	0.6072	0.3543
0.2073	0.6209	0.3925	0.3772	0.7272
0.3270	0.7003	0.3892	0.2211	0.6315
0.4330	0.1120	0.3426	0.6120	0.1918
0.2388	0.0294	0.5087	0.7849	0.7468
0.4133	0.6692	0.7216	0.1697	0.4502
0.6859	0.3936	0.5534	0.5187	0.4502
0.4622	0.4290	0.3320	0.2657	0.3054
0.5697	0.1456	0.4303	0.6336	0.4735

By Equation (15) to identify the cases of nonlinear problems and practical, and effective; the recognition rate of 100%, identification of new classes and the actual output of the sample Figure 3 (Recognition function y=∑i=1lαi yiK (xi,x)+b, The first 24 samples of a class, identify the dialogue, recognize the value of y> b; 21 for the two types of specimens, identification dialogue, recognize the value of y <b), and identify the open side between two types of "zone" That recognition results are obvious. The [1] used five linear classifier, the recognition results were not necessarily intended to do (see, Figure 2), the procedures and algorithm complexity. The methods described here, and put forward for the most effective optimization algorithm, which has good versatility and can be used in many areas.

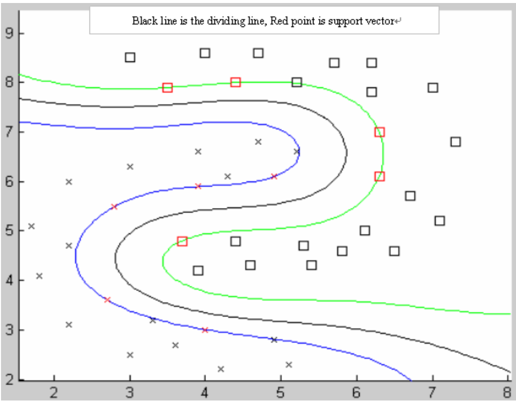

Figure 3. SVM and genetic algorithm for nonlinear pattern recognition results

6. Conclusion

As a high-profile field of machine learning support vector machines SVMs shows great advantages. Support vector machines are based on structural risk minimization principle, to maximize the generalization ability of learning machine, that is, from the limited training samples are small errors on the independent test set is still able to guarantee a small error. Support vector machine algorithm is a convex optimization problem, so the local optimal solution is globally optimal solution. Support vector machine algorithm using the kernel function in an implicit nonlinear transformation, cleverly solved the curse of dimensionality.

The genetic algorithm and support vector machine combined effectively to high-speed nonlinear pattern recognition, it is superior to other methods. And the genetic algorithm has the following main features:

1) The genetic algorithm is coded to operate on the parameters, rather than the parameters themselves;
2) Genetic algorithms start from many points in parallel operation, not just the point;
3) Genetic algorithm to calculate the adaptive value of the objective function without the other is derived, and thus less dependent on the problem;
4) The rules of the genetic algorithm optimization is determined by the probability, not certainty;
5) Genetic algorithm for efficient solution space heuristic search, rather than blindly exhaustive or completely random search;
6) Optimization of genetic algorithm is a function to be the basic unrestricted, it neither requires a continuous function, differentiable function is not required, either expressed by mathematics were analytic functions, and can be even a neural network mapping matrix the implicit function, and thus a wider range of applications;
7) The genetic algorithm has the characteristics of parallel computing, which can be massively parallel computing to increase computing speed;
8) The genetic algorithm is more suitable for large-scale optimization of complex issues;
9) The genetic algorithm is simple, powerful.

References

[1] Leon Bobrowski. Design of piecewise linear classifiers from formal neurons by a basis exchange technique. *Pattern Recognition.* 1991; 24(9): 863-870.
[2] Chang-qun. Wang Xiaolong and support vector classification and multi-width of the Gaussian kernel, *Journal of Electronics*. 2007; 35(3); 484-487.
[3] Abhijit, S Pandya, Robert, B Macy forward. neural network pattern recognition and its implementation. Xu Yong, Jing Tao, translation, Beijing. Publishing House of Electronics Industry. 1999; 6.
[4] Wang Xiaoping, Cao Liming with genetic algorithms - theory, application and software implementation of [M], Xi'an, Xi'an Jiaotong University Press. 2002; 6.
[5] Cortes C, Vapnik V. Support Vector Networks.*Machine Learning.* 1995; 20; 273-295.
[6] V Vapnik. Statistical Learning Theory. New York: Wiley. 1998.
[7] V Vapnik. The Nature of Statistical Learning Theory. New York: Springer-Verlag. 1995.
[8] Wang Peng, Zhu Xiaoyan. RBF kernel based SVM model selection and its application. *Computer Engineering and Applications.* 2003; 39(24): 72 ~ 73.
[9] Li Lin, ZHANG Xiao. SVM based on RBF kernel optimization learning algorithm. *Computer Engineering and Applications.* 2006; 42(29): 190 ~ 192.