Advanced Engineering Forum *Online: 2012-09-26*
ISSN: 2234-991X, Vols. 6-7, pp 694-699
doi:10.4028/www.scientific.net/AEF.6-7.694

SVM Parameter Optimization Based on Immune Memory Clone Strategy and Application in Bus Passenger Flow Counting

Zhu Fang[1, a] , Wei Junfang[2, b]

[1]School of Computer and communication engineering, Northeastern University at Qinhuangdao, Qinhuangdao, 066004, China

[2]School of Resource and material, Northeastern University at Qinhuangdao, Qinhuangdao, 066004, China

[a]sky050607@sina.com, [b]weijunfang123@sina.com

Keywords: support vector machine, parameters selection, immune memory clone, bus passenger counting

Abstract. The performance of support vector machine (SVM) depends on the selection of model parameters, however, the selection of SVM model parameters more depends on the empirical value. According to the above deficiency, this paper proposed a parameters optimization method of support vector machine based on immune memory clone strategy (IMC). This method can solve the multi-peak model parameters selection problem better which is introduced by n-folded cross-verification. Tests on standard datasets show that this method has higher precision and faster optimization speed compared with other four methods. Then the proposed method was applied to bus passenger flow counting. The experimental results show that the method reposed in this paper obtains higher classification accuracy.

Introduction

The Support Vector Machine (SVM) is a new machine learning method that based on the Statistic Learning Theory (SLT) [1,2]. The selection quality of SVM parameters and kernel functions has an effect on the learning and generation performance. In order to find the best parameters for SVM, many researchers have done a mass of study. The parameters in SVM are usually selected by man's experience, such as n-folded cross-verification [3]. Recently, there are some automatic parameter selection methods researched such as colony algorithm and genetic algorithm [4-7]. These methods are efficient and automatic for optimizing parameters in a certain degree. But they depend on optimization model construction, and convergence to local optimum sometimes. According to these problems, a parameters optimization method of SVM based on immune memory clone strategy (IMC) is proposed in this paper. The results of experiment show that the proposed method has more efficiency of optimization and higher accuracy rate of classification than other existent methods.

Parameters Optimization Algorithm of SVM Based on Immune Memory Clone Strategy

Support Vector Machine. SVM is based on the principle of structural risk minimization. The ideal of SVM is to search for an optimal hyperplane to separate the data with maximal margin. When the training set is nonlinear, the training vector x is mapped into a higher dimensional feature space by a nonlinear function $\varphi(x)$,and in the feature space who's dimension maybe infinite construct the optimal classification hyperplane and the classifier's decision function. In order to construct the optimal hyperplane, the following optimization problem must be solved:

$$\begin{aligned} &\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i \\ &s.t.\ \ y_i((\varphi(x_i) * w) + b) \geq 1 - \xi_i,\ \ \xi_i \geq 0, i = 1, \cdots, l \end{aligned} \tag{1}$$

Where * is inner product, w are coefficient vector, $\xi_i \geq 0$,are slack variables and C is a penalty parameter to be chosen by user. Finally, the decision function as follow:

$$f(x)=\mathrm{sgn}\left(\sum_{i=1}^{n}\alpha_i^{*}y_iK(x_i,x)+b^{*}\right) \tag{2}$$

Where x_i are Support Vectors (SVs), Lagrange multipliers α_i satisfy with $0<\alpha_i^{*}<C$, n is number of SVs, b^{*} is bias value. Eq.(2) shows that kernel function and penalty parameter affect the performance of SVM[8].

The Immune Clonal Algorithm. Clonal selection is an artificial immune algorithm that is applied to optimization problems. Affinity proportional reproduction and affinity maturation are two important features of the clonal selection. An antigen selects some cells to obtain their clone. The selection rate of each cell is directly proportional to its affinity with selective antigen. If an antigen has a high affinity, its offspring number will be large. The mutation rate is inversely proportional to its affinity with an antigen. The immune clonal algorithm includes affinity operator, antibody concentration operator, clonal operator, mutation operator, clonal selection operator and so on.

Parameters Optimization Algorithm of Support Vector Machine Based on Immune Memory Clone Strategy (IMC-SVM). This paper introduces the memory mechanism and adaptive mutation operator into immune clonal algorithm. The proposed algorithm can both improve searching optimization speed and ensure global searching ability, so it is much suitable for multi-peak value SVM parameters selection problem.

The key of the proposed algorithm for SVM parameters optimization based on artificial immune algorithm is presented as follows:

1) Coding schemes design of antibody. The design of antibody gene encoding schemes is based on principle of coarse search first then finer search. In this proposed algorithm, the binary coding is adopted. For example, the parameters of SVM with RBF kernel include penalty C and kernel parameter σ. Every antibody is composed by two sections: kernel function parameters σ and penalty C. The range is [1, 10000] and [0.0001, 1] respectively. Increment $\Delta\sigma$ is 1and increment ΔC is 0.0001. Every parameter can be represented by 15 bit binary, so coding scheme of antibody is composed by 30 bit binary.

2) Antibody-antigen affinity. Because this proposed algorithm is inspired by multi-peak value optimization based on n-folded cross-verification, the result of n-folded cross-verification is taken as antibody-antigen affinity so that the unbiased estimate for generalization of SVM can be ensured.

3) Antibody-antibody affinity. The Hamming distance for binary encoding is taken as the norm of antibody-antibody affinity. Its definition is given as follow:

$$aff(ab_i,ab_j)=\sum_{k=0}^{L-1}\alpha_k\ ,\ \alpha_k=\begin{cases}1 & ab_{ik}=ab_{jk}\\ 0 & ab_{ik}\neq ab_{jk}\end{cases} \tag{3}$$

Where ab_{ik} is k position in antibody i, ab_{jk} is k position in antibody j, L is the length of encoding.

4) Clone. The clonal operator is described as follows:

$$Tc(ab_i)=clone(ab_i) \tag{4}$$

Where $clone(ab_i)$ is the set composed of cells cloned by antibody ab_i, m_i is the number of cells cloned which is obtained by adaptive computation. According to the cloning principle, those antibodies with high affinity will clone more cells. The number of cloned cells determined by the following formula:

$$m_i=\mathrm{int}\left[N_c\cdot\frac{aff(a_i)}{\sum_{j=1}^{n}aff(a_j)}\right] \tag{5}$$

Where N_c is the parameter of scale of cloned cells, int() is a function to get the integral part of its variable.

5) Mutation. The memory unit and populations adopt different mutation operator. The mutation operator of memory unit can be described as follows:

$$T_m(ab_{ijm}) = \begin{cases} !ab_{ijm} & rand() < pre \\ ab_{ijm} & rand() \geq pre \end{cases}, \quad pre = \frac{\lambda}{aff(ab_{ijm})BD} \tag{6}$$

Where ab_{ijm} is the j position in L-dimensional of cell m cloned by antibody ab_i, B=j, D is iteration number, $rand()$ is a random function which generate a random number in the range of (0, 1), pm is mutation rate. The mutation operator of populations can be described as follows:

$$T_m(ab_{ijm}) = \begin{cases} g\_best_j & rand() < \inf\_factor \\ !ab_{ijm} & rand < pm \\ ab_{ijm} & else \end{cases} \tag{7}$$

Where g_best is the optimal solution of antibody population, $\inf\_factor \in [0,1]$ is the parameter directly proportional to the degree of close to g_best, $rand()$ is a random function which generate a random number in the range of (0, 1), pm is mutation rate.

6) Clonal selection. The cell with low affinity in memory is instead by antibody with high affinity in population. Those cells with low affinity are mutated according to formula, and then put into population. Meanwhile the antibody with low affinity must be eliminated. In order to preserve the diversity of the antibody, the same number individual is generated and put into population.

$$T_m(ab_{ijm}) = \begin{cases} !ab_{ijm} & rand() < pm \\ ab_{ijm} & else \end{cases} \tag{8}$$

The steps of optimization algorithm are described below.

Step 1: First, the location of approximate region of parameters is determined, and then N encoded antibodies are selected randomly in this region as initial population A. The size of population is selected as 20. Step 2: P antibodies are generated randomly; the affinities between A and P are calculated. S maximum affinity antibodies are selected from A to form memory unit. The others in A form population L. Step 3: The antibody-antigen affinity in memory unit and population L are calculated. If termination condition is met, the algorithm is stopped the optimum solution is found, else go to next step. Step 4: Under optimal selecting ratio, choose n antibodies with the high affinity from antibody population and memory unit to clone to generate the temporary antibody population C. Step 5: Antibodies selected randomly from the cloned cells above are mutated, the antibodies are eliminated, which affinity is lower than its parents. Step 6: The mutated cells are clonal selected, and then go to step 3.

Contrast Experiment on Five Algorithms

In this section, we present some experimental results on testing accuracy and average run time on a suite of four datasets from UCI benchmark repository. All experiments are carried out on a PC machine with CPU-AMD5200and 1G memory under C++6.0 platform and Libsvm package.

The proposed optimization algorithm was performed on SVM with kernel RBF. The parameters include kernel parameter σ and penalty parameter C. The experiment is carried on five times on each dataset. The experimental results are shown as table 1. In table 1, σ and C are optimal parameters, T is runtime of optimization algorithm, R is the optimal result of cross-verification.

As is shown from table 1 above, the speed of IMC-SVM, DE-SVM and PSO-SVM is faster than Grid-search. Moreover, the selected parameters are better and the cross-verification error is less. Compared with DE-SVM and PSO-SVM, the searching optimization speed of ICM-SVM is faster, especial on svmguid3 dataset, and the cross-verification accuracy of ICM-SVM is higher. Compared with GA-SVM, the searching optimization speed of both is close. But the cross-verification accuracy of IMC-SVM is higher than GA-SVM. Aim to multi-peak value optimization problem, IMC has better searching optimization ability than GA. The IMC is more suitable for multi-peak parameter optimization based on cross-verification.

Table 1 Experimental result on SVM with kernel RBF

Dataset	optimization algorithm	C	σ	T/s	R%
	Grid-search	5869	0.04	8.781	94.9839
	PSO-SVM	9007	0.0045	4.935	95.3106
svmguid1	DE-SVM	13099	0.0233	4.759	95.3441
	GA-SVM	13937	0.0089	4.289	95.472
	IMC-SVM	14371	0.0092	4.26	96.1879
	Grid-search	1903	0.327	6.662	79.34
	PSO-SVM	1009	0.096	3.975	80.772
svmguid3	DE-SVM	792	0.762	3.78	81.4815
	GA-SVM	872	0.502	2.33	81.1906
	IMC-SVM	548	0.5305	2.328	82.1983
	Grid-search	7993	0.6	13.09	88.97
	PSO-SVM	10023	0.007	7.548	89.05
splice	DE-SVM	7985	0.05	6.339	89.5172
	GA-SVM	9812	0.0924	5.76	89.024
	IMC-SVM	7005	0.0325	5.789	90.0203
	Grid-search	3089	0.076	7.793	98.995
	PSO-SVM	445	0.734	3.976	99.3089
ijcnn1	DE-SVM	1002	0.982	3.875	99.7963
	GA-SVM	9277	0.3225	3.296	99.076
	IMC-SVM	5977	0.4011	3.081	99.8217

Bus Passenger Flow Counting Method Based on IMC-SVM

According to the current research on the bus passengers flow counting, this paper proposes a new counting method of bus passenger flow based on SVM. This method solves the problem that low cost device can't count the number of passengers accurately and can't distinguish the direction of passengers' movement.

Data preprocessing. We put a pedal on the stair of the bus. Under the four corners of the pedal, there are four pressure sensors for analog output. Four channel analog signals were jointed together, and then connected to the A/D converter of the data collector. The date collector based on Sumsung 2410 arm9 Single Chip Micyoco with A/D converter. Its frequency reaches to 200MHZ. And the sampling rate is 20ms. Then the data will be transferred to PC through RS232 .The system frame diagram is shown in Fig.1.

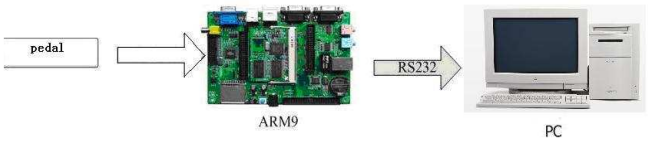

Fig.1 The system frame diagram of data gathering

The experiments proved that the passage time of one people getting on and off the pedal is more than 300ms. So if the passage time is less than 300ms, the corresponding pressure data will be seen as noise. Data preprocessing is necessary before feature extraction and pattern recognition. We use the method of sliding window, which is similar to the algorithm of smoothing processing in the image processing. A linear template of 1*3 is taken as filter. The result proves that this method is effective

Feature extraction. When people go through the pedal the pressure sensor generates continuous analog signal. The analog signal is collected at 20ms intervals. It will be converted into digital signal by the collector. This digital signal can be fitted into a curve after noises removing. We extract the parameters of this curve as a group of characteristic vectors of SVM. These characteristic vectors are as follows:

(1)The value of the first peak (F1) and the time interval from the beginning of the wave to the first peak (T1);

(2) The value of second peak (F2) and the time interval from the second peak to the end of the wave (T2);

(3) The first peak's slope (R1) and the second peak's slope (R2) (the peak's slope is defined as follow: R1=F1/T1, R2=F2/T2);

(4) The difference value between the values of two peaks (F3) (F3=F2-F1);

(5) The ratio of R1 and R2 (R3) (R3=R1/R2);

The parameter sensitivity interval of two-category algorithm based on SVM is [-1, 1]. So the vector processing is necessary to make sure every data is in this interval. Because the max value of the A/D is 1024, every data can be divided by 1100. The vector matrix [F1,T1,R1,F2,T2,R2,R3,F3] is taken as the input characteristic matrix . The output vector is a one-dimensional vector [X], the value of X can be 1 and -1. 1 means getting on and -1 means getting off.

Test

4488 groups of data were obtained through experiments and tested by the SVM and IMC-SVM. The result is shown in the following table. According to the experiment results below, it obviously shows that our algorithm has higher recognition accuracy than the general support vector machines.

Table 2 Recognition of the test samples

Model	test samples	Recognition samples	Recognition rate of SVM (%)	Recognition rate of IMC-SVM (%)
Getting on	2462	2375	94.4663	96.1256
Getting off	2026	1843	90.9674	92.1326
Total	4488	4218	93.9840	94.1291

Conclusion

Based on Immune Memory Clone Strategy, a novel parameter optimization algorithm is proposed. In the process of parameters selecting, combined with cross-verification to ensure the unbiased estimate for generalization. The great global searching ability of Immune clone algorithm is used to realize automatic selection of SVM parameters. The simulation experiments prove that the proposed algorithm has higher accuracy and optimization speed than other four algorithms. Then the proposed method was applied to bus passenger flow counting. The experimental results indicate that this method has higher recognition accuracy. It has a good application foreground.

References

[1] H.G. Zhou, C.D. Yang, Using Immune Algorithm to Optimize Anomaly Detection Based on SVM, in: Proceedings of IEEE International Machine Learning and Cybernetics Conference, Dalian, China, 2006, pp. 4257-4259

[2] J. Bo, T. Yuchun, Z. Yang-Qing, L. Chung-Dar, I. Weber, Support Vector Machine with the Fuzzy Hybrid Kernel for Protein Subcellular Localization Classification, in: Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ'05), Reno, NV, 2005, pp. 420-423

[3] C. Nello, S.T. John: An Introduction to Support Vector Machines and other Kernel-Based Learning Methods, Electronic Industry Press, 2006

[4] Zhao Chunxiu, Zhou Huiren, Liu Chunxia, Application and Parameters Optimization of LS-SVM based on SA and Bootstrap, Statistic and Decision, 2010, pp.25-28

[5] Liu Xiangying, Jiang Huiyan, Tang Fengzhen, Parameters Optimization in SVM Based-on Ant Colony Optimization Algorithm, Nanotechnology and Computer Engineering, 2010, pp. 470-480

[6] Zhu Ning, Feng Zhigang, Wang Qi, Parameter Optimization of Support Vector Machine for Classification Using Niche Genetic Algorithm, Journal of Nanjing University of Science and Technology (Natural Seience), 2009, pp. 16-19

[7] Ren Yuan, Bai Guangchen, Determination of Optimal SVM Parameters by Using GA/PSO, Journal of Computers, 2010, pp. 1160-1165

[8] Liu Hanbing, Jiao Yubo, Damage identification for simply-supported bridge based on SVM optimized by PSO, Information Engineering for Mechanics and Materials, 2011, pp. 490-494