License plate character recognition based on GA-SVM

Weichao Jiao [a], Junfei Dong and Xiufang Yuan

China West Normal University, Nanchong 637000, China

[a]1093938619@qq.com

Abstract

Support vector machine is a better performance of the machine learning algorithm, its parameters has important implications to the accuracy of the classification, select parameters has also become one of the main direction of the current research in machine learning. For parameter selection problem, we propose SVM parameter optimization method based on genetic algorithm, GA-SVM model was constructed and applied to classify identify the license plate characters. Experimental results have shown that: Compared SVM and grid optimization algorithm, GA-SVM algorithm has a higher license plate character recognition accuracy, character recognition accuracy of its training set up to 100%, the character recognition accuracy of the test set to reach 98.98%.

Keywords

Support vector machine; genetic algorithm; parameter optimization; character recognition.

1. Introduction

Automatic license plate recognition system (License Plate Recognition, LPR) is one of the key technologies of intelligent transportation systems (Intelligent Traffic System, ITS) in. License plate recognition system mainly consists of three parts: license plate location, character segmentation, character recognition. Among them, the character recognition is a core part of the license plate recognition system, the accuracy of recognition results will directly affect the performance of the license plate recognition system. At present, the common license plate character recognition methods are based on the template matching method [1-2], on the neural network method [2-3] and so on. Proceed from the overall of the characters, template matching to test with the standard template matching characters, character by calculating their correlation for identification. This method is faster and has a higher recognition accuracy, but the license plate images collected in a natural environment often exist rotation or license plate characters fade, deformation, etc., these can cause serious impact to the recognition results would be severely affected by these recognition results; neural network has a strong curve fitting and pattern classification capabilities, which get it in the character recognition is widely used, but this algorithm has slow convergence and easy to fall into local minimum also other shortcomings.

Support vector machine (support vector machine, SVM), with its small sample, nonlinear, high dimensional pattern recognition features have been more and more widely used in character recognition. Different selected of the parameters will affects the performance of SVM model, the literature [4-5] has analyzed the influence of kernel function parameters g and penalty factor C to SVM performance, but have not given a specific selection method yet. Huang [6-7] and Fei [8] using the PSO (particle swarm optimization, PSO) parameter C and g optimization. In this paper, SVM parameter optimization method based on genetic algorithm (genetic algorithm, GA). Experimental results have shown that: this algorithm relatively SVM, PSO-SVM license plate character recognition with higher accuracy.

2. SVM Fundamentals

SVM is based on statistical learning theory of machine learning methods, it proposed by Vapnik in 1995. It with empirical risk and confidence risk that minimize the sum of the theoretical basis of

structural risk minimization principle, have stronger generalization ability. The purpose of their study was to find a hyper plane, which can accurately classify and make two kinds of sample points to its minimum and maximum distance. Known training set: $T=\{(x_1,y_1),...,(x_l,y_l)\}\in(X\times Y)^l$ Wherein $y_i\in Y=\{1,-1\}\ (i=1,2,...,l)$; x_i is the feature Vector.

Lagrange multiplier method and introduces kernel will look for separating hyper plane problem into:

$$\min_{\alpha}\frac{1}{2}\sum_{i=1}^{j}\sum_{j=1}^{l}y_iy_j\alpha_i\alpha_jK(x_i,x_j)-\sum_{j=1}^{l}\alpha_j$$

$$s.t.\begin{cases}\sum_{i=1}^{l}y_i\alpha_i=0 & i=1,2,...,l\\ 0\le\alpha_i\le C\end{cases} \tag{1}$$

The optimal solution: $\alpha^*=(\alpha_1^*,...,\alpha_l^*)^T$ .Which α_j is Lagrange multiplier; K is kernel function. Select a positive component α_j^* which is located in the open interval $(0,C)$, and according to equation (2) calculate the classification threshold b^*:

$$b^*=y_j-\sum_{i=1}^{l}y_i\alpha_i^*K(x_i-x_j) \tag{2}$$

Decision function is constructed:

$$f(x)=\mathrm{sgn}(\sum_{i=1}^{l}\alpha_i^*y_iK(x,x_i)+b^*) \tag{3}$$

More than two classification algorithms can only solve the problem, the problem for multi-classification SVM to realize the idea of using one, using the cross-validation group k (k-fold cross validation, $k-CV$) ideological training, the data into k groups, followed by a group as a test set, and the rest as a training set, resulting in k models, with the average of these k model classification accuracy as a classifier at $k-CV$ performance.

Support vector machine was commonly used in nuclear function, there are many: linear kernel (Linear kernel), polynomial kernel (Polynomial kernel), radial basis (RBF) kernel function, Sigmoid kernel function and so on. The study has shown that: SVM with RBF kernel to kernel has a strong learning ability and classification results. This selection of RBF kernel:

$$K(x,x_i)=\exp(-g\,\|x-x_i\|^2) \tag{4}$$

3. Genetic algorithm optimization parameters

The main parameter of SVM model contains a margin of error penalty factor C and the kernel function parameter g. The different value of parameter, the classifier performance will also have a greater difference. About optimized SVM parameters selection, now the international community has not recognized the best way to unity, since too large C will occur over the study state that the training and test sets a high classification accuracy rate is very low classification accuracy set, so can verify the highest classification accuracy of all parameters C and g, it will select a smaller C value. Since the main characteristic of GA is not dependent on the implicit gradient information, and the ability to group random search, easy to get global optimal solution. In this paper, GA global search ability parameter optimization, to obtain the optimal parameters C and g.

3.1 Determine the size of the population

Species populations over the General Assembly so that the calculation process is quite complex, species populations are too small and not easy to obtain better results, so the size of the appropriate population has important implications for the convergence of GA. Population size is determined by

the computational complexity of the specific issues, generally select the range $[20,100]$. As used herein, seeking support vector machine parameters for correct classification rate greater impact model calculates the amount is too large, it is not suitable for the establishment of large-scale population, therefore, the population size 20 to a smaller value.

3.2 Select the fitness function

Use GA to select parameters, usually fitness function to evaluate individual performance and then search, the search space of knowledge is rarely used in this process, so the fitness function is an important part of GA. Using the k above set of cross-validation as to verify the performance of the index, selecting the formula (5) fitness function:

$$f = (l_1 + l_2 \cdots + l_k) / l \times 100\% \tag{5}$$

In the formula, the number of samples for testing of samples l_i correctly classified the sample set in which the total number of the samples i . l is the test sample set.

3.3 Select and copy operation

Choices are the results from the previous generation of genetic fitness of selected large individual into the next generation of operating with a certain probability. Therefore, the greater the probability that an individual fitness value is also relatively greater will be selected. Wherein the individual selection probability is:

$$P_i = f_i / \sum_{j=1}^{N} f_j \tag{6}$$

In the formula, f_i represents the fitness of individuals, N represents the population size. The number of i individuals to be copied is $R(i) = N \times P_i$. After selecting and copying, it generated a new sub-group from the initial population.

3.4 Crossover and mutation operation

Crossover operation is a process of the main genetic operators in genetic algorithm, because the parameter value coding gene is real, so, in order to guarantee a new parameter value after crossing, and can generate a new search space, cross process operating parameters using a linear combination of genes the way, the two gene strings corresponding to the value of cross-digit combination to generate two new gene string, so that the new gene sequence containing the characteristics of its parent. Selection, crossover operation is substantially complete genetic algorithm most search functions, and mutation increases the ability of the algorithm to find near optimal solution. The main purpose is to maintain the diversity of variation groups, due to the lower population mutation probability may prevent the loss of important single gene to a certain extent; higher mutation probability of genetic operations tend pure random search, would greatly reduce algorithm stability and convergence rate. According to the actual problem to select mutation probability, the general value interval is $[0.001, 0.5]$.

4. Application and comparison of GA-SVM

4.1 GA-SVM character recognition algorithms

Our license plate character usually consists of 31 characters, 25 uppercase letters of the alphabet A-Z and 10 Arabic numerals 0-9. In this paper, the experiment collected 600 license plate character images as the training set, 200 license plate character images as the testing set, each picture size is 32 × 16 pixels.

By using the strengthen toolbox Libsvm, through Matlab programming to construct support vector machine [9-10], following these steps:

Step 1: input data for image processing and image analysis in accordance with the algorithm steps support vector machine parts shown.

Step 2: parameter initialization. The maximum number of iterations to evolve into M=100; population size N=20; mutation probability P_m =0.01; crossover probability P_c =0.9.

Step 3: binary code, randomly generated N chromosomes as the initial population.

Step 4: For each chromosome decoding, and calculating the fitness of each chromosome in accordance with the formula (5).

Step 5: Select the copy according to individual fitness, generate new population.

Step 6: Population of these new generation of genetic manipulation.

Step 7: determine whether the termination condition is satisfied, if satisfied then jump to Step 8, otherwise Step 4.

Step 8: perform decoding of each chromosome, build support vector machine model, and verify the performance criteria to evaluate the use of support vector machine performance.

Step 9: according to the trained SVM model to classify the recognition of the license plate characters.

4.2 Results and Comparison

Genetic algorithm iterations 50 times the highest significance in the CV classification accuracy was 98.9857%, recognition accuracy training samples of 100%, the fitness curve has been showed in Fig. 1.

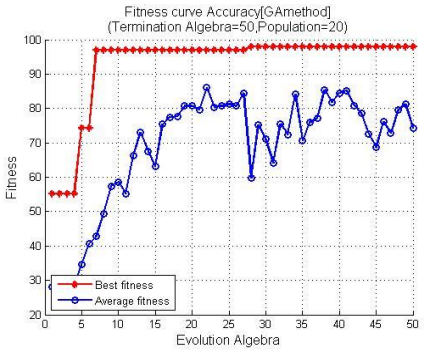

Fig. 1 The fitness curve based on GA parameter selection

The article GA-SVM character recognition algorithm and SVM grid comparative analysis, the results has been showed in Table 1.

Table 1 Comparison of different parameters optimization

Parameter optimization method	Training set recognition rate	Testing set recognition rate
Grid Algorithm	94.2965%	91.3685%
SVM	95.0781%	90.2904%
GA-SVM	100%	98.9847%

We can be seen from Table 1, the angle grid algorithm and SVM, using a higher classification accuracy to identify the genetic algorithm optimization parameters of SVM.

5. Conclusion

This paper puts forward a kind of license plate character recognition algorithm based on GA and SVM, the SVM using GA good optimization performance of one of the important parameters to further improve the classification performance of SVM. On the license plate character data sets of simulation experiment shows that GA - SVM fusion algorithm compared with the basic classification of SVM algorithm and grid algorithm has better performance, so the algorithm is an effective license plate character recognition algorithm. The future direction of the research will be further in-depth

study of the fusion algorithm about the influence of the parameters for the classification results, and further improve the calculation efficiency of the algorithm.

References

[1] W. W. Huang, X. Han, Z. Q. Sen etc. License plate character recognition based on template matching and neural network, Pattern Recognition And Artificial Intelligence, Vol.13(2001) No.01,p.123-127.

[2] W. Min, H. X. Han, W. Wu etc. License plate character recognition method for template matching and neural network. Journal Of Huazhong University of Science And Technology (Natural Science Edition) Vol. 29(2001) No.03,p.48-50.

[3] G. R. Yan, H. X. Hui. Application of BP neural network in vehicle license plate character recognition , Computer Simulation, Vol.27(2010)No.09,p. 299-301+350.

[4] Z. X. Xia, Q. Feng. Choose Gaussian kernel SVM and model parameters, Computer Engineering And Applications, Vol. 42(2006) No.01, p.77-79.

[5] Z. X. Yun, L. Y. Cai. Performance Gaussian kernel support vector machine, Computer Engineering, Vol. 29(2003) No.08, p. 22-25.

[6] L. C. Huang, F. J. Dun. A distributed PSO-SVM hybrid system with feature selection and parameter optimization, Applied Soft Computing, Vol.8(2008) No.04,p.1381-1391.

[7] Y. L. Pan, J. Du, W. T. Hu. Air to ground target optimal tracking method by ant colony optimization-based SVM. Journal Of Computational Information Systems,Vol.10(2014) No.05,p.1805-1810.

[8] F. W. Sheng, W. J. Ming, M. B. Yu etc. Particle swarm optimization based support vector machine for forecasting dissolved gases content in power transformer oil, Energy Conversion And Management,Vol.50(2009) No.06,p.1604-1609.

[9] C. H. Wu, G. H. T zeng, Y. J. Gao etc. A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy,Expert Systems With Applications, Vol. 32(2007) No.02,p.397-408.

[10] S. H. Min, J. Lee, I. Han. Hybrid genetic algorithms and support vector machines for bankruptcy prediction, Expert Systems With Applications, Vol. 31(2006) No.06, p.652-660.