

A New Multi-class Classification Method Based On Least Square Support Vector Regression Machine

Zhaoqing Song¹

The 7th Department of Naval Aeronautical and Astronautical University, Yantai, 264001, China

E-mail: szqln@163.com

Yao Chen²

First team of graduate students' brigade, Naval Aeronautical and Astronautical University, Yantai, 264001, China

E-mail: rt_cy@163.com

Zhenkai Guo³

School of Mathematics & Information, Ludong University, Yantai, 264001, China

E-mail: zhkguo@126.com

The support vector machine is a commonly used classification method for its good performance. The processes of most of the existing methods for multi-class problem are not simple. More than one support vector machine (SVM) classifier should be trained in each of these methods. In this paper, a novel multi-class support vector machine method is presented, named the Multi-class Least Square Support Vector Regression (MCLS-SVR) by using only one classifier to solve the multi-class problem. The main idea of this method is to transform the classification problem to the regression problem, which is much simpler, by treating the class label of each sample as the regression output value; then LS-SVR is used to solve the regression problem and a rounding operation is added in the course of testing samples because their class labels are positive integers. The results of both simulation and experimental on face and fingerprint recognition show that the proposed MCLS-SVR method yields higher recognition rate and faster calculation speed.

CENet2015

12-13 September 2015

Shanghai, China

¹Speaker

²Corresponding Author

³The foundations that have support this paper are: National Natural Science Foundation of China (No.: 61433011), and the General Program of Shanghai Postdoctoral Research Project (12R21414300).

1. Introduction

The support vector machine (SVM) is one of the most popular classification methods because of its better data generalization performance and being able to solve nonlinear, small sample size and high-dimension problems. Since SVM was proposed by Vapnik[1], a lot of researches have been carried out to apply and improve this method.

SVM was originally developed for binary classification problems, but in practice, there are usually more than two classes of samples in most of the classification problems; so we should extend the binary SVM to multi-class SVM. There are mainly two types of approaches now:

One approach is to decompose the classification problems into several binary SVM classifiers and then reconstruct them. As to the first approach, there are many methods: One-Against-Rest (OAR)[2], One-Against-One (OAO)[3], Directed Acyclic Graph (DAG) SVM[4], etc. These methods all have some deficiencies because they failed to consider the classification problem as an entirety; however, lots of researches have been carried out to improve these methods and overcome their deficiencies, such as: Shiladitya Chowdhury et al., Henry Joutsijoki, etc.[5-6].

The other approach is to solve the MC problem with only one SVM classifier[7-8], so that the information of all samples would be kept; however, it would not be easy to design the SVM classifier and the calculation speed may be slow if the sample size is large.

This paper presents a new multi-class classification method belonging to the second approach. This method is called Multi-class Least Square Support Vector Regression which introduces the idea of regression into classification problem. This method considers all the data and contains only one classifier which is easy to train. The experimental results show that this method performs very well and features a high classification rate.

The organization of the paper is shown as follows: Section 2 introduces the support vector machine theories; Section 3 presents the proposed the LS-SVR based multi-class classification method; Section 4 gives the experimental results and Section 5 makes the conclusion

2. Support Vector Machine Theories

In Section 2, we give a brief description of SVM, SVR and LS-SVR methods.

2.1 Support Vector Machine

The Support vector machine is a non-probabilistic supervised learning model[9]. Consider a classification problem, the training data set of which is shown as follows:

$$S_1 = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\}, \quad \mathbf{x}_i \in R^m, y_i \in \{1, -1\} \quad (2.1)$$

where \mathbf{x}_i are the input vectors, y_i are the class labels of \mathbf{x}_i and l is the size of training data set, $i = 1, 2, \dots, l$.

As to the linear separable problem, the goal of SVM method is to find an optimal hyperplane by solving a quadratic programming (QP) problem :

$$\begin{aligned} \min & \frac{1}{2} \|\boldsymbol{\omega}\|^2 + C \left(\sum_{i=1}^l \xi_i \right) \\ \text{s.t.} & y_i (\boldsymbol{\omega} \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \dots, l \end{aligned} \quad (2.2)$$

where $\boldsymbol{\omega}$ is the normal vector of the hyperplane; C is the penalty factor; ξ_i are the slack variables. Lagrange function is introduced to solve the QP problem and the optimal classification function is as follows :

$$f(\mathbf{x}) = \text{sgn} \left[\sum_{i=1}^l y_i a_i^* (\mathbf{x}, \mathbf{x}_i) + b^* \right] \quad (2.3)$$

where $\mathbf{a}^* = (a_1^*, a_2^*, a_3^*, \dots, a_l^*)$ are the Lagrange multipliers, b^* is the classification threshold, and the minority of samples which the nonzero Lagrange multipliers are corresponding to the called Support Vectors (SV).

As to the nonlinear classification problems, the kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ is introduced and the optimal classification function changes to:[10]

$$f(\mathbf{x}) = \text{sgn} \left[\sum_{i=1}^l y_i a_i^* K(\mathbf{x}, \mathbf{x}_i) + b^* \right] . \quad (2.4)$$

There are many commonly-used kernel functions and we choose Gaussian radial basis function (RBF) kernel: $K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / 2\sigma^2)$ in this paper.

2.2 Support Vector Regression Machine

If we introduce an alternative loss function to the SVM method, it can also solve regression problems[11]. Consider the same training data set as Equation (2.1), the goal of the regression problem is to find a function $f(\mathbf{x})$ to express the relationship between \mathbf{x}_i and y_i . According to the Structural Risk Minimization (SRM) principle, $f(\mathbf{x})$ should minimize following formula,

$$R = \frac{1}{2} \|\boldsymbol{\omega}\|^2 + CR_{emp} , \quad (2.5)$$

where R is the expected risk, R_{emp} is the empirical risk and C is penalty factor. There are different kinds of R_{emp} functions, and the ε -Insensitive Loss Function is widely used because it usually uses fewer support vectors[12]. The function is defined as follows:

$$|y - f(\mathbf{x})|_{\varepsilon} = \begin{cases} 0 & |y - f(\boldsymbol{\omega}, \mathbf{x})| \leq \varepsilon \\ |y - f(\boldsymbol{\omega}, \mathbf{x})| - \varepsilon & \text{Otherwise} \end{cases} , \quad (2.6)$$

where ε is the given accuracy. It shows that we don't have to care about errors less than ε ; especially, when $R_{emp} = 0$ and the slack variables is added, the optimization problem is changed to:

$$\begin{aligned} \min & \frac{1}{2} \|\boldsymbol{\omega}\|^2 + C \left(\sum_{i=1}^l \xi_i + \xi_i^* \right), \\ \text{s.t.} & y - f(\mathbf{x}_i) \leq \varepsilon + \xi_i, \xi_i \geq 0 \\ & f(\mathbf{x}_i) - y \leq \varepsilon + \xi_i^*, \xi_i^* \geq 0 \end{aligned} \quad (2.7)$$

Upon solving Equation (2.7) by the Lagrange multiplier method, we can get the following regression function:

$$f(\mathbf{x}) = \sum_{SVs} \hat{a}_i K(\mathbf{x}, \mathbf{x}_i) + b \quad (2.8)$$

where $\hat{a}_i = (a_i - a_i^*)$ are Lagrange multipliers and $K(\mathbf{x}_i, \mathbf{x})$ is kernel function.

2.3 Least Square Support Vector Regression Machine

Suykens J. A. K. and Vandewalle J. has proposed an improved method for SVM named Least Square Support Vector Regression (LS-SVR). In this method, Equation (2.7) has been transformed into a group of equality constraints; thus the solving process of SVM has been changed from QP problem to some linear equations, which has reduced the difficulty of training and raised the solution efficiency significantly.

As to the same training data set as Equation (2.1), the goal of LS-SVR is to solve the following problem:

$$\begin{aligned} \min & \frac{1}{2} \|\boldsymbol{\omega}\|^2 + \frac{C}{2} \left(\sum_{i=1}^l e_i^2 \right) \\ \text{s.t.} & y_i = \boldsymbol{\omega} \boldsymbol{\varphi}(\mathbf{x}_i) + b + e_i, i = 1, \dots, l \end{aligned} \quad (2.9)$$

where $\mathbf{e} = [e_1 \ e_2 \ \dots \ e_l]^T$. This problem can be solved by Lagrange multiplier method, so the regression function is

$$f(\mathbf{x}) = \sum_{i=1}^l a_i K(\mathbf{x}, \mathbf{x}_i) + b \quad (2.10)$$

where a_i is Lagrange multiplier.

3. Multi-class Least Square Support Vector Regression

In this section, we propose a LS-SVR based multi-class classification method. Compare a N classes classification problem with a regression problem, we can see that: in the classification problem, the category labels $y_i \in \{1, 2, \dots, N\}$ are all positive integers; in the regression problem, the outputs $y_i \in \mathbb{R}$ are real numbers; as a result, if we can make conversion between regression outputs and category labels, the multi-class classification problem can be treated as a regression problem, which is more easily to be solved.

Given a N -class classification training sample set:

$$S_1 = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\}, \quad \mathbf{x}_i \in \mathbb{R}^m, y_i \in \{1, 2, \dots, N\}, \quad (3.1)$$

We use the LS-SVR method to train these samples and get Equation (2.10) as the classifier. To classify an unknown sample, we apply it into Equation (2.10) and round $f(\mathbf{x})$ to get the final classification result, that is:

$$f(\mathbf{x}) = \text{round} \left[\sum_{i=1}^l a_i K(\mathbf{x}, \mathbf{x}_i) + b \right], \quad (3.2)$$

where $\text{round}(x)$ is defined as follows:

$$\text{round}(x) = \begin{cases} p, & p - 0.5 \leq x \leq p + 0.5 \\ p - 1, & p - 1.5 \leq x \leq p - 0.5 \end{cases}, \quad (3.3)$$

where p is positive integer and $p \geq 2$.

With this function, multi-class classification problem can be directly solved by the regression method, which is simple and fast; the robustness of the method is greatly enhanced with the introduction of $\text{round}(x)$, and we can still classify correctly when there are noises in the input data.

Essentially, this method has omitted the intermediate process of the classification and sought the mapping relationship between the input samples and their class labels. This method has greatly simplified the training process and accelerated the speed of operation by solving the classification problem with the idea of regression. Next, we should verify the validity of this method.

4. Experimental Results

In order to test the proposed method, we have selected three datasets from the UCI database: Iris dataset, Wine dataset and Image segmentation dataset. The information of the datasets is shown in **Table 1**. To train the LS-SVR classifiers, we randomly selected 80% of samples from each dataset for training, and the rest for testing; and we chose the RBF kernel $K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / 2\sigma^2)$. Let $\sigma = 30$, $C = 2000$, then we calculated the decision functions (Equation (12)) respectively. The testing samples are used for verification of the method.

As a comparison, we solve these multi-class classification problems with the OAO method at the same time, and the simulation results of classification accuracy rates are shown in **Table 2**. It could be seen that Multi-class LS-SVR method performed better than the OAO method and the classification accuracy rates were very high.

	Number of instances	Number of attributes	Number of classes
Iris	150	4	3
Image segmentation	2310	19	7
Wine	178	12	3

Table 1: Information of the Three Datasets

	Iris	Wine	Image segmentation
MCLS-SVR	100%	93%	100%
OAO	97%	86%	98%

Table 2: Classification Accuracy Rates of the Two Methods

We then applied the method to the face recognition and the fingerprint recognition to verify its effectiveness. Some of the face samples were selected from the Olivetti Research Laboratory (ORL) database and others were created by us. Each image contains the face region and the non-face region as well as change of expression and illumination, including 80 faces images from 20 persons (4 images each). The fingerprint samples are collected by us too, including 60 fingerprints from 30 persons (2 images of each).

For higher recognition rate, the samples need to be pretreated before feature extraction. The pretreatment process in this paper is as follows:

- Gray processing for the original color images;
- Select feature regions from the images, shear on the gray images appropriately and normalize them;
- Filter the normalized images;
- Edge detection on the images;
- Binary processing to the images.

Then the invariant moment[13] is used for the image feature extraction upon these pretreatment.

The step of face recognition by using Multi-class LS-SVR method is as follows:

Firstly, take three samples from each category to form the training samples, the rest as the testing samples; then train the LS-SVR with training samples to get the decision function Equation (3.2); finally randomly select a sample from the testing sample set and introduce it into the decision function and compare the output of decision function with its class label. The classification is successful if they are equal.

Following the step above, we picked up 60 face images from 20 persons to form the training set. We chose the RBF kernel $K(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2)$, let $\sigma = 64$, changed parameter C to see its influence on results, and the result is shown in **Table 3**. The process of fingerprint recognition was basically the same with face recognition, 60 images were divided into training set and testing set, each contained 30 images. We let the parameter $\sigma = 40$ and the test result is shown in **Table 4**.

Experimental results of face recognition and fingerprint recognition showed that the Multi-class LS-SVR method also performed very well in practical application. We could notice that the parameter C should be large enough to get higher classification rate.

C	10	100	1000	100000
Recognition rate	35%	65%	95%	100%

Table 3: Recognition Percentage of Faces Images

C	10	100	1000
Recognition rate	56.67%	96.67%	100.00%

Table 4: Recognition Percentage of fingerprint Images

5. Conclusion

This paper presents a novel scheme for multi-class classification problem by using Multi-class Least Square Support Vector Regression, the main idea of which is to convert the classification problem to the regression problem, which is more easily to solve. The difference between LS-SVM and LS-SVR lies in their output values: the integer for LS-SVM and the real number for LS-SVR; thus we add a rounding operation to SVR so that its output values can correspond to the class labels we need. Thus only one LS-SVR needs to be trained to solve the MC classification problem. The Simulation and application results show that the proposed MCLS-SVR performs well and has higher classification rate.

References

- [1] V. Vapnik, *The nature of statistical learning theory*[M]. New York: Springer-Verlag, 25-27(1995).
- [2] U. Kressel, *Pairwise Classification and Support Vector Machines*[J]. Advances in Kernel Methods: Support Vector Learning, 255-268(1999).
- [3] P. Bennett Kristin, *Combining Support Vector and Mathematical Programming Methods for Classification*[M]. USA: MIT Press, 307-326(1999).
- [4] C. Platt John, Cristianini Nello, *Shawe-Taylor John. Large Margin DAGs for Multiclass Classification*[C]. Neural Information Processing Systems, 547-553(2000).
- [5] C. Shiladitya, S. J. Kanta, Basu Dipak Kumar, et al. *Weighted multi-class Support Vector Machine for robust face recognition*[C]. International Conference on Communications, Devices and Intelligent Systems, 326-329(2012).
- [6] H. Joutsijoki. *Half-Against-Half Multi-Class Support Vector Machines in classification of benthic macroinvertebrate images*[C]. International Conference on Computer & Information Science, 414-419(1012).
- [7] A.-G. Jeronimo, P.-C. Fernando. *Multi-class support vector machines: a new approach*[C]. International Conference on Acoustics, Speech, and Signal Processing, 6-10(2003).
- [8] W. Jason, W. Chris. *Support vector machines for multi-class pattern recognition*[C]. The European Symposium on Artificial Neural Networks, 219-224(1999).
- [9] C. Corinna, V. Vladimir. *Support-vector networks*[J]. Machine learning, 20(3): 273-297(1995).
- [10] A. Ma, E. M. Braverman, Rozonoer Li. *Theoretical foundations of the potential function method in pattern recognition*[J]. Automation and Remote Control, 25(6): 917-936(1964).
- [11] J. Smola Alex, *Regression estimation with support vector learning machines*[D]. Physik Department, Technische Universität München, (2003).
- [12] C. Vladimir, M. Filip. *Learning from data: concepts, theory, and methods*[M]. New York, NY, USA: John Wiley & Sons, 35-40(1998).

- [13] C. Gonzalez Rafael, *Digital image processing*[M]. Boston, MA, USA: Addison-Wesley Longman Publishing, 1992: 484-486.

Pos (CENet2015) 032