



Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

An image classification algorithm using fuzzy support vector machine

Cao Jianfang^{1,2}, Chen Junjie² and Chen Lichao³

¹Department of Computer Science & Technology, Xinzhou Teachers University, Xinzhou City, China

²College of Computer & Software, Taiyuan University of Technology, Taiyuan City, China

³College of Computer Science & Technology, Taiyuan University of Science and Technology, Taiyuan City, China

ABSTRACT

The development of electronic technology and imaging technology has resulted in the rapid growth of digital images. It has become an urgent problem to rely on advanced technology to identify images. An image recognition algorithm based on fuzzy support vector machine is proposed. The algorithm makes up for the lack of traditional support vector machine in multi-classification problems and solves the problem of semantic ambiguity in image classification by defining fuzzy membership function. Using 6 types of natural images to test, the experimental results show that classification performance improves significantly compared with the traditional support vector machine method.

Key words: fuzzy support vector machine, fuzzy membership function, features extraction, image semantics, image classification

INTRODUCTION

With the development of electronic technology and the Internet technology, and the rapid popularization of image acquisition devices such as digital cameras, camera device and electronic camera etc. in recent years, the digital image has a rapid expansion of the scale and there will be GB even TB of digital images to be generated, published and shared every day. Images play a very important role in people's life. People can perceive rich information containing in the image when they observe an image: the target, the spatial layout, function and scene semantic, even the emotion semantic image brings to people^[1,2]. However, the massive image data brings people not only convenience, but also a lot of problems: People are difficult to find the information of their real needs in the voluminous data and information. Faced with such a large-scale image resources, how to make the organization and classification effectively for them has become an urgent problem to be solved at present.

Currently, there are many algorithms to be used for image classification, such as the Bayes classifier, K-nearest neighbor classification, neural network, genetic algorithm, clustering etc. But only under the condition of adequate samples, can the performance of these classification methods have the theoretical guarantee, and in fact, the number of samples is often limited. Therefore the classification effect of the above algorithms in practical application is not just as one wish.

Support vector machine (SVM), proposed by Vapnik, is a new machine learning method, and is based on VC dimension theory and structural risk minimum principle of statistical learning theory. It finds the best compromise between model complexity and learning ability according to the limited sample information in order to get the best generalization ability^[3,9,10]. In the field of image classification, SVM has the advantages of small sample learning, nonlinear, high dimension, anti-noise performance, high learning efficiency and the good generalization, etc.

This paper is organised as follows. In Section 2 we first discuss the fuzzy support vector machine algorithm. Then, in Section 3, we propose an algorithm for image classification based on fuzzy support vector machine in detail. Then, experimental results are presented in Section 4. Finally, Section 5 presents our conclusions and future work.

II. FUZZY SUPPORT VECTOR MACHINE

A. Standard support vector machine

The core idea of the standard support vector machine is to create a low VC dimension optimal classification hyperplane in high dimensional kernel space by putting learning samples mapping into a high dimensional kernel space nonlinear. The classification function of which the upper bound of risk is minimized is obtained, considering the size of experience risk and confidence interval, according to the principle of structural risk minimization for its eclectic.

Supposing separable sample set is $(x_i, y_i), i = 1, \dots, n, x \in R^n, y \in \{+1, -1\}$, the main purpose of SVM is to construct a hyperplane to split the two kinds of different samples, which makes the classification interval maximum. Thus we have the following quadratic optimization problem:

$$\begin{aligned} \min & \left(\frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \varepsilon_i \right) \\ \text{s.t.} & \quad y_i [\omega x_i + b] - 1 + \varepsilon_i \geq 0 \\ & \quad \varepsilon_i \geq 0 \end{aligned} \quad (1)$$

Introducing Lagrange multipliers, we get its dual form:

$$\begin{aligned} \max & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s.t.} & \quad 0 \leq \alpha_i \leq C \\ & \quad \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned} \quad (2)$$

Among them, ε_i is the slack variables, C is the penalty coefficient, $K(x_i, x_j)$ is the kernel function. Thus the decision function is:

$$f(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right) \quad (3)$$

B. Multi-class support vector machine

SVM applied to multi-class classification problem can be roughly divided into two categories^[3].

(1) One-to-one. The method is to train a SVM classifier in each of the two classes of samples, which can obtain $k(k-1)/2$ classifiers, and the number of classifier associated with each category is $k-1$. When predicting the unknown sample, each classifier judges its category and the corresponding category wins a vote. The category of unknown sample is the final category having the most votes.

(2) One-to-many. The method is to use a SVM classifier to distinguish each class separately from other types of class, resulting in the k classifiers. When predicting the unknown sample, the sample is classified in that category with the greatest decision function values.

For the above two methods, when the number of classes is more, training speed and classification speed is slower and they have no-identifiable areas. Then there are several improved algorithm, including ECOC, DAG etc. For ECOC method, it's different to select appropriate samples; For DAG method, its shortcomings are: (i) When the category number for the classification is much, the training speed is low; (ii) the selection of root node may directly affect the result of classification.

C. Fuzzy support vector machine

(1) Introduction of fuzzy support vector machine

Support vector machine divides the samples into two mutually antagonistic classes through optimal hyperplane. However, in the practical application, each sample is not completely can be grouped into a class in some cases, that is to say, there are some fuzzy membership between samples and classes. Thus many scholars introduces the fuzzy theory into the support vector machine, and put forward the fuzzy support vector machine in order to solve some problems existing in the traditional support vector machine, which is the further improvement and perfection of the traditional support vector machine. The basic idea is: The fuzzy technology is applied to the support vector machine, and the different samples adapt to the different punishments coefficient, which results in different samples have

different contributions when the objective function is constructed. And samples with noise or wild values get less weight so as to achieve the purpose of eliminating noise and affection of outliers in the sample ^[10].

When using fuzzy support vector machine for classification, it's different from the conventional SVM for the representation of the training sample that fuzzy support vector machine adds membership in addition to the characteristics of the sample and generic identification. Supposing the training sample set is expressed as:

$$(x_i, y_i, \mu(x_i)), i = 1, 2, \dots, n, x \in R^n, y \in \{+1, -1\}$$

Among them, $\mu(x_i)$ is a membership, $0 < \mu(x_i) \leq 1$. Since the membership $\mu(x_i)$ indicates the degree that the sample belongs to a certain class and ε_i is classification error term in the objective function of SVM, $\mu(x_i)\varepsilon_i$ is the error term with the weight, and the optimal classification plane we get is the the optimal solution of the following objective function:

$$\begin{aligned} \min & \left[\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \mu(x_i) \varepsilon_i \right] \\ \text{s.t.} & \quad y_i [wx_i + b] - 1 + \varepsilon_i \geq 0 \\ & \quad \varepsilon_i \geq 0 \end{aligned} \quad (4)$$

Adding Lagrange multipliers $\alpha_i (i = 1, 2, \dots, n)$, we get its dual form:

$$\begin{aligned} \max & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s.t.} & \quad 0 \leq \alpha_i \leq C \mu(x_i) \\ & \quad \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned} \quad (5)$$

The decision function is:

$$f(x) = \text{sign} \left[\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right] \quad (6)$$

By comparing the formula (5) with (2), we can see that fuzzy support vector machine (FSVM) is different from support vector machine (SVM) on the constraint conditions. The parameter C in SVM is a custom penalty factor and controls the degree of punishment for wrong classified samples. The larger C is, the greater the punishment is and the degree of constraint for wrong classified samples, the smaller the interval of the classification is. With the loss of C , support vector machine algorithm would ignore more samples and get a classification surface with a larger edge interval. For fuzzy support vector machine, if we set the value of C larger and take the membership degree $\mu(x_i)$ with 1, fuzzy support vector machine degrades general support vector machine, which allows smaller classification error rate and gets classification surface with narrower edge. The fuzzy support vector machine algorithm can reach the effect of different degree of punishment with different samples through giving different membership degree for different samples. The smaller the membership degree is, the smaller role in training sample is. As a result, the fuzzy support vector machine has stronger anti-interference ability than support vector machine.

(2) The determination of fuzzy membership degree

When using fuzzy technology to process, the design of the membership function is the key for algorithm. Different membership functions has different impact on processing results of algorithm and the implementation of algorithm, which requires membership function must be able to objectively and accurately reflect the uncertainty about the sample in the system. At present, there are a lot of methods to construct membership functions, but there are not general guidelines to follow. When dealing with actual problems, we usually need determine reasonable membership function aiming at specific problems and according to the experiences. Many scholars made some researches in this area, but currently the method of using the distance between the sample and class center to measure the size of its membership in fuzzy support vector machine methods. The method because of fewer amounts of calculation and strong robustness gets more widely recognized.

This paper will use the measure method of fuzzy membership degree to determine the fuzzy membership $\mu(x_i)$. It regards the membership as a function of the distance between the samples in feature space and its class center. Let x_0 be the class center, r be class radius, r is determined by the following formula.

$$r = \max_i \|x_i - x_0\| \quad (7)$$

Then the membership of each sample is:

$$\mu(x_i) = 1 - \frac{\|x_i - x_0\|}{r} + \delta \quad (8)$$

Wherein, δ is a small preset constant ($\delta > 0$) to avoid the situation $\mu(x_i) = 0$.

III. IMAGE CLASSIFICATION USING FUZZY SUPPORT VECTOR MACHINE

A. The general process of image classification

The general process of image classification is: image capture, image preprocessing, image feature extraction, classifier decision, etc. The classification process is as follows in figure 1.

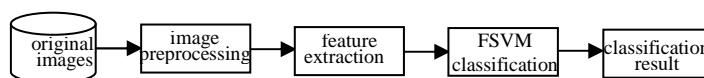


Figure 1. The process of image classification

B. Image preprocessing

Because the actual image data are always incomplete, have noises and are inconsistent, preprocessing becomes more and more important. There are two kinds of data preprocessing technique for data mining of image data^[4-5]: data cleaning and data transformation. The first step of image preprocessing is to use the denoising technology to process images. The denoising of images may remove most of the background information and noise. Classification effect also depends on the features extracted from the image to a large extent. The second step is image enhancement in image preprocessing. In the process of image generation, transmission or transformation, image quality always reduces due to the influence of various factors. The purpose of image enhancement is to adopt a series of technology to improve the effect of the image or transform the image into a more suitable form for processing. We can use the widespreadly used histogram equalization technology to realize the image enhancement.

C. Feature Extraction

Feature extraction is a crucial step in the process of image classification. The quality of feature extraction will directly affect the performance of the classifier^[6-8]. This paper uses three kinds of characteristics (color, texture, edge direction) to express image content for an image.

(1) Color feature

The colors of image carry a wealth of semantic information; color feature is most commonly used visual features in image classification. RGB color, being composed of R , G , B components, is the most common color space. Now a variety of images use RGB space to store and transport, and RGB color gets direct support of various physical devices. However, the studies find that the RGB components are not independent of each other, and exist some correlation. But HSV color space has a better independence property of visual perception, and it can feel continuously change from all kinds of color components. In the other word, the color change that the human eye can perceive can be measured by the Euclidean distance of the color components. Therefore this paper adopts HSV color space. Calculation process is as follows: given color point (r, g, b) of the RGB color space, the color point (h, s, v) of the HSV color space is:

Assume: $v' = \max(r, g, b)$, define: r' , g' , b'

$$ENG = \sum_i^G \sum_j^G P^2(i, j | d, \theta)$$

$$h' = \begin{cases} 5+b' & \text{if } r = \max(r, g, b), g = \min(r, g, b) \\ 1-g' & \text{if } r = \max(r, g, b), g \neq \min(r, g, b) \\ 1+r' & \text{if } g = \max(r, g, b), b = \min(r, g, b) \\ 3-b' & \text{if } g = \max(r, g, b), b \neq \min(r, g, b) \\ 3+g' & \text{if } b = \max(r, g, b), r = \min(r, g, b) \\ 5-r' & \text{otherwise} \end{cases}$$

$$r' = \frac{v-r}{v-\min(r, g, b)}$$

$$g' = \frac{v-g}{v-\min(r, g, b)}$$

$$b' = \frac{v-b}{v-\min(r, g, b)}$$

$$h = 60 \times h' \quad s = \frac{v-\min(r, g, b)}{v} \quad v = \frac{v'}{255}$$

Among them,

$$r, g, b \in [0, 255], h \in [0, 360], s \in [0, 1], s \in [0, 1].$$

For HSV color space, the distance between $c_1 = (h_1, s_1, v_1)$ and $c_2 = (h_2, s_2, v_2)$ can be calculated using the following formula:

$$d(c_1, c_2) = [(v_1 - v_2)^2 + (s_1 \cosh_1 - s_2 \cosh_2)^2 + (s_1 \sinh_1 - s_2 \sinh_2)^2]^{1/2} \quad (10)$$

The similarity between two colors is:

$$s(c_1, c_2) = 1 - \frac{1}{\sqrt{5}} d(c_1, c_2) \quad (11)$$

(2) Texture feature

Texture reflects local structure feature of the image, specifically say, it is change of the pixel grayscale or color in the image of a pixel neighborhood. The change is related in the spatial statistics, which consists of texton and arrangement of primitive. This paper proposes texture feature extraction method based on proposed GLCM combining the definition of texture. GLCM reflects comprehensive information of image gray on direction, interval and magnitude of changes based on second-order joint conditional probability density function between image gray degrees^[11,12]. Supposing $P(i, j | d, \theta)$ is the probability of occurrence of each pixel having a gray level i and j giving space distance d and direction θ , which is usually expressed in matrix form, called GLCM. If the gray level of an image is G , the co-occurrence matrix is a $G \times G$ square matrix. Image texture features can be described by calculating secondary statistic of GLCM. This method defines 14 texture features, only four of which are not relevant: energy, homogeneity, contrast and correlation. The computing methods are as follows.

Energy:

$$ENG = \sum_i^G \sum_j^G P^2(i, j | d, \theta) \quad (12)$$

Homogeneity:

$$IDM = \sum_{i=1}^G \sum_{j=1}^G P(i, j | d, \theta) / [1 + (i - j)^2] \quad (13)$$

Contrast:

$$CON = \sum_{i=1}^G \sum_{j=1}^G (i - j)^2 P(i, j | d, \theta) \quad (14)$$

Correlation:

(15)

$$\begin{aligned}\mu_1 &= \sum_{i=1}^G i \sum_{j=1}^G P(i, j | d, \theta) \\ \mu_2 &= \sum_{j=1}^G j \sum_{i=1}^G P(i, j | d, \theta) \quad \sigma_1^2 = \sum_{i=1}^G (i - \mu_1)^2 \sum_{j=1}^G P(i, j | d, \theta) \\ \sigma_2^2 &= \sum_{j=1}^G (j - \mu_2)^2 \sum_{i=1}^G P(i, j | d, \theta)\end{aligned}$$

Energy is local homogeneity measure of images. For the homogeneous regions, the particular pixel pairs are more and few element values of GLCM are larger, which result in that the energy values are larger. Homogeneity reflects the degree of texture thickness and gray uniformity, high-value array elements which concentrate near the main diagonal and the values are larger. Contrast measures the amount of local change and reflects the degree of the groove depth of the texture. In formula (14), $|i - j|$ indicates gray-scale difference of the pixel pairs. The more the pixel pairs with large gray-scale difference are, the greater the contrast value is. Correlation measures the degree of similarity of GLCM elements in direction of row or column and reflects the correlation between pixel gray values of images.

(3) Edge feature

Edge feature is the most basic characteristics because the edge includes the boundary information of valuable target which can be used for image analysis, image filtering, and object recognition. Canny operator is used to test the edge in this paper and we can get a binary image retaining only the edge, and then calculate the following characteristics for the edge image.

(i) Edge length. That is the number of the non-zero pixels in the image. The value indicates the total length of the inner edge of the image.

(ii) Minkowski dimension, namely Minkowski fractal dimension of edge. This feature can effectively describe the extent of the sparse of image edges. Calculation of the fractal dimension needs use a series of structural elements of different sizes to expand the edge image. With the increase of the size of the structural elements, the total number of non-zero pixels (edge points) is also increasing. The more intensive image edge is, the faster speed of the increasing of total number N of edge points with the enlargement of size S of structure element is. We make linear fitting for logarithm of N and S , and get the slope G of the fitted line. The fractal dimension of the image is defined $D=2-G$.

D. Classification method

According to the classification process shown in the above figure 1, input the original image, the first step is to make pretreatments, such as noise reduction, enhancement etc. This step has the ability of reducing speckle noise in flat areas and keeps details with obvious changes. Then, extract color, texture and edge features. RGB color space is converted into HSV color space when extracting color feature. Texture features includes energy, homogeneity, contrast and correlation. Edge feature includes edge length and Minkowski dimension. Finally, the extracted feature vector is input to the FSVM classifier and we can obtain classification results.

RESULTS AND DISCUSSION

Experimental data come from ordinary natural images from Internet. These images are divided into six classes: architecture, animal, landscape, plant, people, interior decorating. We use support vector machine and fuzzy support vector machine for classification respectively. Sample images used in the experiment are shown in Table 1.

TABLE 1: IMAGE SAMPLES USED IN THE EXPERIMENT

sample classification	architecture	animal	landscape	plant	people	interior decorating
Train sample	200	200	160	180	160	120
Test sample	170	180	160	150	150	100

The indexes defined in this paper to characterize classification performance of support vector machine and fuzzy support vector machine in the multi-class classification problems are:

The number of unclassified samples *USS*: the number of samples in test set being not classified (ie. The sample is neither properly classified nor misclassified).

Unclassified rate *USR*: the proportion of unclassified samples in total samples in the test set.

$$USR = \frac{USS}{N_t} \quad (16)$$

N_t is total number of samples in test set.

The number of classified samples *CSS*: the number of samples which are not properly classified by SVM, but are properly classified by FSVM.

Classified sample rate *CSR*: the proportion of classified samples in unclassified samples.

$$CSR = \frac{CSS}{USS} \quad (17)$$

The classification results of the two methods are shown in Figure 2 and Table 2.

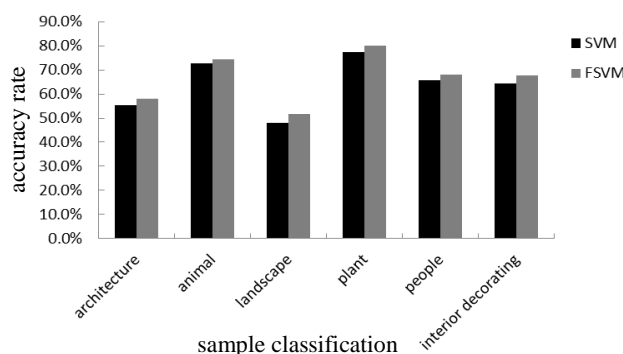


Figure 2. Comparison of classification rate between SVM and FSVM

TABLE 2 : COMPARISON OF CLASSIFICATION PERFORMANCE BETWEEN SVM AND FSVM

sample classification	the number of test samples	USS	USR	CSS	CSR
architecture	170	8	4.3%	4	52.0%
animal	180	12	7.1%	5	42.1%
landscape	160	5	2.9%	0	0.0%
plant	150	7	5.0%	3	43.4%
people	150	4	2.8%	0	0.0%
interior decorating	100	6	6.2%	2	34.1%

We can clearly see from the above experimental data that classification performance of fuzzy support vector machine is superior to that of support vector machine. The degree of superior depends on the number of not classified samples. The more the number of not classified samples, the better the performance of fuzzy support vector machine is. If not having not classified samples, fuzzy support vector machine has the same performance as support vector machine.

CONCLUSION

This paper uses fuzzy support vector machine to make the classification research on natural images. Aiming at the problem of no-identifiable areas existing in the support vector machine when dealing with multi-class classification problem, we introduce fuzzy membership function to solve the problem. In order to evaluate the effect and performance of classification methods, we have done experiments using different types of image. The experimental results show that image classification method based on fuzzy support vector machine is superior to that based on support vector machine.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China under Grant No. 61202163 and by the Natural Science Foundation of Shanxi Province under Grant No.2012011011-5 and No. 2013011017-2 and by the Technology Innovation Project of Shanxi Province under Grant No. 2013150 and by the Key disciplines supported

by Xinzhou Teachers University under Grant No. XK201308. The authors are grateful for the constructive and valuable comments made by the many expert reviewers.

REFERENCES

- [1] Jen sen J R. *Introductory Digital Image Processing: A Remote Sensing Perspective*. Third Edition. *Beijing: China Machine Press*, **2007**.
- [2] J. Li, Allison N M. *Neurocomputing*, vol.71, no. 10-12, **2008**, pp.1771-1787.
- [3] W. Ying, Z. G. Wang, J. L. An. *Computer Engineering*, no.16, **2006**, pp. 74-76.
- [4] Marin F, Nozha B. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no.4, **2007**, pp. 818-826.
- [5] X. Sun, Y. D. Xie, D. C. Ren, *Computer Engineering and Design*, vol. 31, no. 21, **2010**, pp.4653-4654.
- [6] G. Y. Wang, J. F. Kang. *Computer Engineering and Design*, vol. 32, no. 1, **2011**, pp. 236-239.
- [7] Rachel LO T W, Paul Siebert J. *Computer Vision and Image Understanding*, vol. 113, no. 12, **2009**, pp. 1235-1250.
- [8] Chow T W S, Rahman M K M. *Neurocomputing*, vol. 70, no. 4-6, **2007**, pp. 1040-1050.
- [9] Hall M, Frank E, Holmes G, et al. The WEKA data mining software: an update// *Sigkdd Explorations*. *New York, NY, USA: ACM*, **2009**, pp.10-18.
- [10] TOMMASI T, ORABONA F, CAPUTO B. An SVM confidence-based approach to medical image annotation // *Proc of Cross-Language Evaluation Forum*. *Berlin: Springer-Verlag*, **2008**, pp.696-703.
- [11] F. Xie, L. T. Chen, H. Qiu. *Application Research of Computers*, vol. 26, no. 7, **2009**, pp. 2767-2770.
- [12] H. X. Sun, Y. H. Zhang, F. L. Luo. *Journal of Optoelectronics Laser*, vol. 21, no. 3, **2010**, pp. 448-451.