Hyperspectral Image Classification Based on Compound Kernels of Support Vector Machine

Yuyong Cui *, Zhiyuan Zeng and Bitao Fu

Digital Engineering and Simulation Research Center, Huazhong University of Science and Technology, Wuhan 430074, China

Abstract. Support vector machine is a kind of pattern classification algorithm based on the statistics learning theory. This paper proposes to estimate abundances from hyperspectral image using probability outputs of support vector machines (SVM), training a SVM with a gauss kernel function, and we discussed the relationship between kernel functions and nonlinear mappings and mapped spaces. Then, a new kernel - compound kernel function is given. We applied the kernel in SVM and compared the kernel with other kernels in Hyperspectral image classification, and give a comparison with some result to present the validation in remote sensing, the result shows the evidence to the validity of the method and shows that this method is more accurate than the other method, which has a more accurate result.

Keywords: support vector machine, compound kernels, hyperspectral image, classification, accuracy

1. Introduction

An important use of earth observation satellite technology is hyperspectral image classification. Implicit in the traditional classification process is the concept that each feature vector should be mapped into one of the classes of interest. Now, a theory based on the so-called Structural Risk Minimization (SRM) principle is the Support Vector Machine, this method has been developed by Vapnik and his co-workers[1-2]. SVM is to find the “optimal” hyperplane which separates without errors the training set[10-11], and maximizes the distance, named “margin”, between the points of both classes and the hyperplane. It can be gained preferable applications in pattern recognition through SVM training algorithm, such as image recognition and handwritten digit recognition[3-4].

SVM in remote sensing image classification has much value to improve the classification’s precision. Now it uses in many aspects. In this paper, we describe how support vector machines (SVM) can be used for remote sensing image classification. SVM perform optimal discrimination. In which the data are correctly classified, next the distance from the decision boundary to the classes is mixed. First, we use the gauss kernel to classify the image and then discuss the relationship between kernel functions and nonlinear mappings and mapped spaces. At last, we give a new kernel - compound kernel function, to classify the image, which has a higher precision than the gauss kernel function. And an example illustrates how the SVM based on compound kernels can be used to classify the Landsat TM data set.

2. Support Vector Machine

2.1. Overview of Support Vector Machines

Support vector machines are classification and regression algorithms formulated from Statistical Learning Theory[1,11]. This framework links the learning (empirical) performance of an algorithm to the true performance which is used in practice. The rate of convergence of the empirical estimate to the true performance is a function of the algorithm’s VC-dimension, which is a measure of flexibility. By minimizing

* Yuyong CUI. Tel.: +86 13407106310. E-mail address: cyy_1999@126.com
the model’s flexibility during the learning process the risk of over-fitting the training set is reduced \[11\].

Non-linear classification and regression SVM have been proved more validity than Artificial Neural Networks (ANN), and the empirical performance of SVM is generally as good as the best ANN solution \[5\]. And when classification is more than 2 classifiers, we must construct other SVM, and the SVM should choose the weights for the first layer to be the training input vectors as this minimizes the VC-dimension. A convenient method to deal with the problem is to choose a kernel function. In general, a non-linear SVM can be expressed as follows:

\[ f(x) = \sum_i a_i y_i K(\bar{x}_i, \bar{x}) - b \]  

(1)

Where \( f(x) \) is the output of the SVM, \( K \) is a kernel function which measures the similarity of a stored training example \( \bar{x}_i \) to the input \( \bar{x} \), \( y \in \{-1,1\} \) is the desired output of the classifier which is a threshold that can be distinguish one class to another classes, and \( a_i \) are blending the different kernels\[6\] weights. The training of SVM consists of finding the \( a_i \) and to choose some good sample, the training is shown as minimization of a dual quadratic form:

\[
\text{Minimize} W(\alpha) = \frac{1}{2} \sum_{i,j} a_i a_j y_i y_j K(\bar{x}_i, \bar{x}_j) - \sum_{i=1}^{n} a_i
\]

(2)

\[ s.t. \quad \sum_{i=1}^{n} a_i y_i = 0 \]

\[ C \geq a_i \geq 0 \quad i = 1, \cdots, n \]

\( a_i \) is the Lagrange multipliers of a primal quadratic programming(QP) problem, and there is a one-to-one correspondence between each \( a_i \) and each training example \( x_i \), and \( K(\bar{x}_i, \bar{x}) \) is kernel function which generally has three kinds of types as follows\[7,9,11\].

Standard kernels are polynomials of degree \( d \):

\[ K(\bar{x}, \bar{x}_i) = [(\bar{x}, \bar{x}_i) + 1]^d \]  

(3)

gauss kernels:

\[ K(\bar{x}, \bar{x}_i) = e^{-|\bar{x} - \bar{x}_i|^2 / 2\sigma^2} \]  

(4)

Sigmoid kernels:

\[ K(\bar{x}, \bar{x}_i) = \tanh(k(\bar{x}, \bar{x}_i) + v) \]  

(5)

While, \( k>0, v<0 \)

\[ \text{Fig. 1: The curve of Gauss kernel function} \]
2.2. The Multi-class Problem of SVM Classifier

Firstly, SVM is proposed to solve binary pattern recognition problem, now, there are so many use of SVM to analysis the multi-class which is expanded from two-class case. Generally multi-class SVM classifier has two cases as follows [7,11]:

- several two-class classifiers together is to solve the multi-class problem.
- multi-classification-plane are set to do the multi-classifiers which solve problem into one optimization problem, and the optimization problem is used to finish multi-class pattern recognition problem.

The second method seems to be simple, but it needs to solve large-scale optimization problem, whose accuracy and classification time are not superior to the first method [11]. The first method always can be expressed some types: one-versus-one, one-versus-rest and many-versus-many. These methods are more complex and hard to get a good result, and some methods exits mixed and missed samples. The paper uses the one-versus-rest combined type, and furthermore aims at its disadvantage and uses component kernels to reclassify these mixed and missed samples produced via SVM classifier.

2.3. Kernels function of SVM

The kernel function is important to the SVM. We can see the ability of classifier. Generally, a kernel function \(K: R^n \times R^n \to R\) computes the inner product between two examples value \(K(x, y) = \langle \Phi(x), \Phi(y) \rangle\), where \(\Phi\) is a mapping from the input space to a transformed feature space. In the way, a constrained quadratic optimization problem aimed to maximize the margin or distance of opposite examples to the hyper plane, and to minimize a regularization factor that allows for misclassifications.

Each kernel function has its advantage and disadvantage. The behaviour of kernel is different, and the SVM constructed by the kernel is different, too. So to choose a kernel is to choose a SVM. Now the kernel has two different classifications: the global kernel and local one. the global kernel allows two far points to infect its value, but the local kernel has local attribute, and it only allows very near point to give the influence to the kernel function.

The learning capacity and generalization capacity determine the learning model good or bad. When you determine a kernel, its interpolating theory also determines the SVM’S learning capacity and generalization capacity. Polynomial kernel is a kernel function that satisfies the Mercer condition and easy to generalize. Another kernel function is conditional positive kernel, which bases on the Regular theory, and now it has been demonstrated that can be used for kernel study [8,12]. Positive kernel definition: if a symmetric function \(k: X \times X \to R\) for all \(N \in \mathbb{N}\), \(x_i \in X\), the Gram positive result in a matrix, that is, all the \(c_i \in R\), we get this function:

\[
\sum_{i,j=1}^{N} c_i c_j K_{i,j} \geq 0
\]  

(6)

Where \(K_{ij} = k(x_i, x_j)\),and \(k\) is named positive kernel. Now we give a conditional positive kernel function, and it is also a local kernel function. It doesn’t satisfy the Mercer condition, but it can be used in the learning study. It is as follows:

\[
k(x, y) = -\|x - y\|^q + 1, 0 \leq q \leq 2
\]  

(7)

The conditional positive kernel function is a local kernel function. It makes local data points to impact the close data, only to the local test data points near the Role. It has better interpolation with kernel capability, and superior ability polynomial interpolation kernel, but generalization capacity is limited. Therefore, in order to build a better interpolation capacity, better learning ability to a generalization model, we can make full use of these two categories of the Kernel.

Since the merits of the above combination of the two kernel functions, in order to make the Kernel has not only a good learning ability, but also has the ability to better generalize. According to a Kernel conditions, the two kernel functions and are still eligible to the kernel function [8], this paper makes the following composition of a new kernel:

\[
K(x, y) = \rho_1 k(x, y) - \rho_2 \|x - y\|^q + 1
\]  

(8)
Where $\rho_1$ and $\rho_2$ are the percentage generally, $\rho_1+\rho_2=1$, $k(x,y)$ is one of kernel function which has better generalization capacity, showed in figure 2. This paper give $k(x,y)$ is equation (4), which is showed in figure 1. in the next part, we will use these two kernel functions to train the data.

![Figure 2: The curve of conditional positive kernel](image)

![Figure 3: The curve of compound kernel function](image)

3. Result And Analysis

3.1. Remote sensing image classification using SVM classifier

There is an experimental image sampled from satellite TM (shown in figure 4), which cover a city area in china (as shown in figure 3). According to analytical result by optimum index method, band 3, band 4 and band 5 are used to identify land-use categories. On the other hand, the data of band 3 data easily distinguish water and land area, and we specify a blue color for them. The data of band 4 can show the plants and can be indicate it as green in the figure. And we knows that band 5 has a high reflection rate, which may reflect the bare land and city area, it is indicated by red, at last, the size of image which we study area of 500×400
pixels is chosen.

![Image](image_url)

**Fig. 4: TM(R:5 band G: 4 band B:3 band) image**

According to the experience, analyzing the image visually, it can be divided in 5 categories, such as road (thereinafter as A class), city area (thereinafter as B class), greenland (thereinafter as C class), and for Yangtze River water and lake water have high different reflection, they show different attribute in the remote sensing image. So we define Yangtze River as D class, and E is the lake river, we choose 150 training samples and 700 test samples of each class, and then train and classify them.

The paper first to use the Gauss kernels to train the data, and then use the compound kernel to train the data, experiment suggest that the Gauss kernels can get a more appropriate result than other kernels, during the training, we first to construct five Gauss kernels SVM classifier, then train and adjust sample set again until gaining satisfactory training set and appropriate SVM parameters. At last we use the 5 certain SVM classifier to test the 3500 test samples, and work out the missed and mixed samples, and then compute the classification performance shown in figure 1. and we can find some mixed samples in the classifier. We put these mixed samples to a process of re-classification, it will greatly improve the classification accuracy.

### 3.2. Classification in component kernel of SVM

After constructing the SVM classifier with the Gauss kernel, we can find the kernel function important to the SVM classifier, now, we will construct the SVM classifier with the component kernel. We will train the sample set and appropriate SVM parameters, to compare with the Gauss kernel we set the SVM parameters is the same as last use in the classifier. We get the result in figure two. To improve the result classification, we can use the same method to re class the mixed samples.

### 3.3. Experimental Results Analysis

From table 1, table 2, we can find the precision of SVM classifier which use the Gauss kernel is 94.2%, but we find the precision of the SVM with the component kernel reach 96.3%, which show that method combination of different kernel functions can improve the classification precision. From the above results, we can get the result: if we want to improve the classification precision, we can use the component kernel, and to train the sample set to get a good classification, so combination different kernels is a kind of effective classification method. We also find in the table, the class B and class C classification is not so good, because the mixed pixels exist, and the remote sensing accuracy is a problem, and it is impacted by many reasons, so it has some uncertain reasons and infect the accuracy of the remote image classification, because the road and city are cross each other, so it has a low accuracy. But we can use some theory to improve the accuracy of the classification.
Table 1 the test result of SVM of Gauss Kernel classifier

<table>
<thead>
<tr>
<th>Result</th>
<th>Test samples</th>
<th>CLASS 1 A</th>
<th>CLASS 2 B</th>
<th>CLASS 3 C</th>
<th>CLASS 4 D</th>
<th>CLASS 5 E</th>
<th>Mixed and Missed samples</th>
<th>Recognition rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>A class 700</td>
<td>673</td>
<td>5</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>27</td>
<td>96.1%</td>
<td></td>
</tr>
<tr>
<td>B class 700</td>
<td>19</td>
<td>613</td>
<td>56</td>
<td>1</td>
<td>2</td>
<td>87</td>
<td>87.6%</td>
<td></td>
</tr>
<tr>
<td>C class 700</td>
<td>15</td>
<td>21</td>
<td>630</td>
<td>0</td>
<td>0</td>
<td>70</td>
<td>90%</td>
<td></td>
</tr>
<tr>
<td>D class 700</td>
<td>3</td>
<td>16</td>
<td>5</td>
<td>686</td>
<td>0</td>
<td>14</td>
<td>98%</td>
<td></td>
</tr>
<tr>
<td>E class 700</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>695</td>
<td>5</td>
<td>99.3%</td>
<td></td>
</tr>
<tr>
<td>ALL 3500</td>
<td>710</td>
<td>695</td>
<td>727</td>
<td>688</td>
<td>697</td>
<td>203</td>
<td>94.2%</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 the test result of SVM of component Kernel classifier

<table>
<thead>
<tr>
<th>Result</th>
<th>Test samples</th>
<th>CLASS 1 A</th>
<th>CLASS 2 B</th>
<th>CLASS 3 C</th>
<th>CLASS 4 D</th>
<th>CLASS 5 E</th>
<th>Mixed and Missed samples</th>
<th>Recognition rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>A class 700</td>
<td>682</td>
<td>8</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>97.4%</td>
<td></td>
</tr>
<tr>
<td>B class 700</td>
<td>5</td>
<td>650</td>
<td>45</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>92.9%</td>
<td></td>
</tr>
<tr>
<td>C class 700</td>
<td>0</td>
<td>19</td>
<td>662</td>
<td>6</td>
<td>0</td>
<td>13</td>
<td>94.6%</td>
<td></td>
</tr>
<tr>
<td>D class 700</td>
<td>0</td>
<td>8</td>
<td>2</td>
<td>690</td>
<td>0</td>
<td>0</td>
<td>98.6%</td>
<td></td>
</tr>
<tr>
<td>E class 700</td>
<td>0</td>
<td>6</td>
<td>9</td>
<td>0</td>
<td>685</td>
<td>0</td>
<td>97.8%</td>
<td></td>
</tr>
<tr>
<td>ALL 3500</td>
<td>687</td>
<td>691</td>
<td>722</td>
<td>696</td>
<td>685</td>
<td>17</td>
<td>96.3%</td>
<td></td>
</tr>
</tbody>
</table>

4. Conclusion

In this paper the kernel function is analyzed and its generalization capacity and learning ability. Then give the use in the remote sensing image classification. Because the remote sense has so many factors to consider, and the image is affected by many reasons, it can’t absolutely show the observation object. The kernel study is a very important portion in the SVM theory. We discuss the kernel and give the kernel disadvantage of existing some mixed and missed samples for SVM multi-classifier, we give the component kernel to improve the SVM learning ability, and we use it to classifier sample set, and give a analysis to the result. Experimental results show that the component kernel is superior to the single kernel. In the future, we will do some work to improve the component kernel construction and do some research in the use of component kernel.

5. Acknowledgements

Thanks to all people who helped me in the paper writing, especially, I am so grateful to say thanks to Yifei Ma, Lin Lu, Liu Liu, Mu Zhou. During the time, they gave me so much help, and Mu Zhou helped to train the data. They are my best friends and best co-workers.

6. References