Hyperspectral Image Recognition Using SVM Combined Deep Learning

Yifan Li¹, Junbao Li², Jeng-Shyang Pan¹,³,⁴

¹ College of Computer Science and Technology, Shenzhen Graduate School, Harbin Institute of Technology, China
² Department of Electronics and Information Engineering, Harbin Institute of Technology, China
³ College of Computer Science and Engineering, Shandong University of Technology, China
⁴ School of Information Science and Engineering, Fujian University of Technology, China
lyf3f@163.com, junbaolihit@126.com, jspan@cc.kuas.edu.tw

Abstract

In this paper, we present the combination of deep learning and Support Vector Machine (SVM) applied on the recognition of hyperspectral images. Hyperspectral image recognition is an essential problem in the practical hyperspectral imagery system. While deep learning is capable of reproducing feature vectors with great dimensions out of original data, it leads to great time cost and the Hughes phenomenon. Such nonlinear problem is regarded as obstacles and kernel method appears to be a promising way to solve it. The performance of kernel-based learning system is influenced by the choices of kernel function and parameter greatly. We present the kernel learning method termed Support Vector Machine (SVM) applied on feature vectors supplied by deep learning upon hyperspectral image. The learning system is improved by adjusting the parameters and kernel functions to the data structure for improving performance on solving complex tasks. Experimental results validate the feasibility of the proposed methods.

Keywords: Deep learning, SVM, Hyperspectral image

1 Introduction

In recent decades, hyperspectral imagers have exhibited great potential capability in remote sensing including target detection technology and spectral imaging technology. Among all hyperspectral applications, recognition various land coverings have been one of our essential concerns, which benefits from the great amount of spatial and spectral information, compared to other images. However, such amount of high spectral dimension leads to high dimension feature vectors which create obstacles in applying traditional image classification algorithms [1]. The Hughes phenomenon may come into being during classification procedures if training samples are outnumbered which happens a lot [2].

The main procedure of the hyperspectral image recognition can be divided into two parts: extracting essential features from enormous bands and designing suitable classifiers for significant classification accuracies. Unfortunately, the huge data size of the hyperspectral image is not only bad for detecting valuable information but also increases the difficulties of classifiers construction. These problems, coupled with other disadvantages such as the overfitting of classifiers caused by the noise, will seriously influence the classification performance [2].

In general, there are three categories of HSI spectral-spatial classifiers. First, many spectral-spatial classifications extract spatial and spectral features from HSI before performing classification. Spatial features based on morphological filters [3-6] are widely used in HSI classification; for example, Ghamisi et al. exploit spatial information using extended multi-attribute profiles (EMAPs) [7]. In order to use spatial features, some researches extract spatial features and spectral features, then use spatial and spectral information in a concatenation strategy; However, all those spatial features are handcrafts, which demanded human knowledge. Furthermore, more features mean higher dimensionality and make HSI classification a more time-consuming task [8].

Second, some spectral-spatial classifications take spatial information into the classifier during classification. Simultaneous orthogonal matching pursuit (SOMP) and simultaneous subspace pursuit (SSP) [9] incorporate the spatial correlation between neighboring samples through a classifier based on joint sparsity representation. This kind of methods gives neighboring samples the right of decision and can improve classification [8].

Third, several classification methods attempt to use spatial dependencies after classification in a decision rule or by spatial regularization. Tarabalka [10] proposed a spectral-spatial classification scheme based on the pixel-wise SVM classification, followed by
majority voting within the watershed regions. All those spectral-spatial classifications significantly improve classification results and can be used in succession [8].

Deep learning using neural networks have presented state-of-the-art performances in a wide range of tasks. A great many of the tasks use the softmax activation function for classification. Support vector machine is an widely used alternative to softmax for classification. Using SVMs in combination with convolutional nets have been proposed in the past as part of a multistage process. In particular, a deep convolutional net is first trained using supervised/unsupervised objectives to learn good invariant hidden latent representations. The corresponding hidden variables of data samples are then treated as input and fed into SVMs.

In this paper, we provide a solution to object classification on hyperspectral image using SVM combined deep learning, the following content is divided into 5 parts: related works; proposed method; experiment and result; and conclusion.

2 Related Works

Recently, fully-connected and convolutional neural networks have been trained to achieve state-of-the-art performance on a wide variety of tasks such as speech recognition, image classification, natural language processing, and bioinformatics. For classification tasks, most of these “deep learning” models employ the softmax activation function for prediction and minimize cross-entropy loss [11]. Support vector machine is an widely used alternative to softmax for classification [12]. Using SVMs (especially linear) in combination with convolutional nets have been proposed in the past as part of a multistage process. In particular, a deep convolutional net is first trained using supervised/unsupervised objectives to learn good invariant hidden latent representations [11]. The corresponding hidden variables of data samples are then treated as input and fed into linear (or kernel) SVMs [13]. Other papers have also proposed similar models but with joint training of weights at lower layers [14].

SVM is a kernel-based classifier consisting in projecting data in a high dimensionality space by means of non-linear mapping function $\Phi$ and aiming at determining the optimal separator hyperplane by margin maximization. SVM is a supervised classifier. It has been proposed first for binary classification [15]. In Yichuan Tang’s paper, they demonstrate a small but consistent advantage of replacing the soft-max layer with a linear support vector machine [11]. In Rafika Ben Salem’s work, they proposed method exploited the performance of support vector machine (SVM) in the processing of data with high dimensionality. In their work, they employed all spectral information and two different spatial features which are Extended Multi-Attribute Profile (EMAP) and the mean of neighborhood pixels. Camps-Valls and al. [16] have investigated the combination of spectral and spatial kernels to get an accurate classification of hyperspectral images. Fauvel and al. [17] have registered a significant improvement according to the characterization of each pixel by a stacked vector that concatenate spectral and contextual information extracted by Morphological Profiles (MP). Kang and al. [18] have exploited the performance of edge-preserving filters to develop an accurate spectral-spatial classification outperforming the classification without filtering. Huang and al. [19] have proposed a multi-feature model aiming at constructing a SVM set combining multiple spectral and spatial features.

The spectrum data in database is collected in advance, so it has inconsistency between the spectrums with the data collection. The inconsistency can be considered the nonlinear changing. The relationships of between spectral curves are the classical nonlinear relationship. So the classification is the nonlinear and complex classification problem. Traditional classification methods are not effective to hyperspectral sensing data, among these machine learning methods kernel learning is a feasible and effective nonlinear classifier methods on hyperspectral sensing data. Kernel-based machine learning, which solves the problem of linear learning using kernel trick, is one important preprocessing method of hyperspectral sensing data. Recently, many linear methods are kernelized. For example, Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) [20-22] are kernelized to Kernel PCA (KPCA) and Kernel Discriminant Analysis (KDA) [23], respectively. The performances of these linear methods are improved because the kernel method is used to characterize the complicated nonlinear relationships. Researchers are developing the kernel learning methods. Baudat and Anouar [24], Liang and Shi [25], Lu [26], Chen [27] and Wang [28] developed a series of improved KDA methods. Some other researchers presented the alternative framework of KLPP to develop a framework of KPCA+LPP [29-31] for image recognition, radar target recognition and other researchers improved LPP with kernels [32-36]. Researchers optimized the parameters of kernel function to improve kernel-based learning [27, 33, 37]. These methods select the optimal kernel parameter from a set of discrete values, but the geometry structure of data distribution in the kernel-based mapping space is not changed. Xiong proposed a data-dependent kernel machine learning [34] and Amari presented the support vector machine classifier by modifying the kernel function [20]. In the previous works [27, 34], the authors presented data-dependent kernel for face recognition. Multiple kernel learning methods are developed to solve the kernel model selection problems [38-41]. In the work of Lingping Kong [42-43], their study in the Hierarchical network and Genetic Algorithm have been inspiring for the
design of our own neural network and our experiment.

3 Proposed Method

3.1 Motivation and Contribution

Deep learning based methods are widely used in machine learning. Deep learning learns hierarchical representation, and the higher layer represents increasingly abstract concepts and is increasingly invariant to transformations and scales [8].

For classification problems using deep learning techniques, it is standard to use the softmax or 1-of-K encoding at the top. For example, given 10 possible classes, the softmax layer has 10 nodes denoted by \( p_i \), where \( i = 1, \ldots, 10 \). \( p_i \) specifies a discrete probability distribution, therefore, \( \sum_{i} p_i = 1 \).

Let \( h \) be the activation of the penultimate layer nodes, \( W \) is the weight connecting the penultimate layer to the softmax layer, the total input into a softmax layer, given by \( \alpha \), is

\[
\alpha_i = \sum_{k} h_{ik} W_{kj}
\]  

(1)

then we have

\[
p_i = \frac{\exp(\alpha_i)}{\sum_{j} \exp(\alpha_j)}
\]  

(2)

The predicted class \( i \) would be

\[
i = \arg \max_{i} p_i = \arg \max_{i} \alpha_i
\]  

(3)

To achieve a better result, we fused the SVM method as an advanced classifier. SVM Methods are supervised learning models with associated learning algorithms that analyse data and recognize patterns and are used for classification and regression analysis. The basic SVM takes a set of input features and predicts, are used softmax layer objective to learn the lower level parameters. For each given input, the possible class form, making it a nonprobabilistic binary linear classifier.

Given a training dataset \( \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \) where \( x_i \in R \) and \( y_i \) is either 1 or \(-1\) indicating the class to which the point \( x_i \) belongs, let \( x = [x_1, x_2, \ldots, x_n]^T \). The construction of the hyperplane for a linearly separable problem is \( w^T x + b = 0 \), where \( w \) is the normal vector to the hyperplane and the parameter \( b/\|w\| \) determines the offset of the hyperplane from the origin along the normal vector \( w \). Thus, the margin between the hyperplane and the nearest point is maximized and can be posed as the following problem:

\[
\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i
\]  

(4)

Subject to \( y_i (w^T x_i + b) \geq 1 - \xi_i, i = 1, 2, \ldots, n, \xi_i \geq 0 \)

where \( C \) is a user-defined constant as the penalty parameter of the error term.

In machine learning, one-class classification, also known as unary classification, tries to identify objects of a specific class amongst all objects, by learning from a training set containing only the objects of that class. This is different from and more difficult than the traditional classification problem, which tries to distinguish between two or more classes with the training set containing objects from all the classes.

One-class SVM is an unsupervised algorithm that learns a decision function for novelty detection: classifying new data as similar or different to the training set.

Most deep learning methods for classification using fully connected layers and convolutional layers have used softmax layer objective to learn the lower level parameters. In this paper, we substitute softmax with single-class SVM as classifier. Single-class SVM takes the output of the third layer of the backpropogate procedure of deep learning as input which are vectors with 2000 dimensions. It is proved that SVM combined deep learning (DLSVM) is capable of recognize materials better than using deep learning only.

3.2 Steps and Procedure

Based on our workflow design in Figure 1, our first procedure is to delete some noise band in order to improve the accuracy. Secondly, in virtue of the lack of pixels of some materials, we eliminated two kinds of materials from our experiment.

Finishing the above steps, we have obtained sufficient high-quality data to begin recognition. Firstly, we utilize deep learning transforming original data vectors into vectors of 2000 dimentions; then we divide the vectors of each material into training samples and test samples according to a certain ratio; finally, the vectors are imported into single-class SVM for recognition and we can achieve its accuracy.

3.3 Discussion

Through such procedure, we managed to take advantage of the neural network to obtain ten times the amount of feature value than original ones while combining it with SVM for its ability of classification. The previous process dig into the hidden pattern of different materials making it easier to conduct classification at the risk of Hughes danger which can be avoided by applying SVM that makes high dimensional vector classification possible.
4 Experiment and Result

4.1 Experiment Sighting and Database

Indian Pines dataset is collected under various spectral and spatial resolutions. The spectral curves denote the different remote sensing environments with Airborne platform. The data cube have 224 bands of spectral resolution through 0.4-2.5μm range and it has the spatial resolution of 20m per pixel. We removed the noisy and water-vapor absorption bands and 200 bands of images are used in the experiments. The whole scene consists of 145×145 pixels and 16 classes of interested objects with the size ranging from 20 to 2468 pixels, 9 classes are used in the experiments. One example is shown in Figure 2.

The hyperspectral unmixing algorithms proposed in this work have been tested using the public domain Indian Pines hyper-spectral dataset which has been previously used in many different studies. This image was obtained from the AVIRIS imaging spectrometer at Northern Indiana on June 12, 1992 from a NASA ER2 flight at high altitude with ground pixel resolution of 17 meters. The dataset comprises 145*145 pixels and 220 bands of sensor radiance without atmospheric correction. It contains two thirds of agriculture, and one third of forest, two highways, aril lane and some houses. Ground truth determines sixteen different classes. Water absorption bands (104-108, 150-163 and 220) were removed, obtaining a 200 band spectrum at each pixel. Figure 3 indicates the distribution of different kinds of materials on Indian Pines and Table 1. records the amount of samples of different materials.
Table 1. Groundtruth classes for the Indian Pines and their sample numbers

<table>
<thead>
<tr>
<th>Num</th>
<th>Class</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alfalfa</td>
<td>46</td>
</tr>
<tr>
<td>2</td>
<td>Corn-notill</td>
<td>1428</td>
</tr>
<tr>
<td>3</td>
<td>Corn-mintill</td>
<td>830</td>
</tr>
<tr>
<td>4</td>
<td>Corn</td>
<td>237</td>
</tr>
<tr>
<td>5</td>
<td>Grass-pasture</td>
<td>483</td>
</tr>
<tr>
<td>6</td>
<td>Grass-trees</td>
<td>730</td>
</tr>
<tr>
<td>7</td>
<td>Grass-pasture-mowed</td>
<td>28</td>
</tr>
<tr>
<td>8</td>
<td>Hay-windrowed</td>
<td>478</td>
</tr>
<tr>
<td>9</td>
<td>Oats</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>Soybean-notill</td>
<td>972</td>
</tr>
<tr>
<td>11</td>
<td>Soybean-mintill</td>
<td>2455</td>
</tr>
<tr>
<td>12</td>
<td>Soybean-clean</td>
<td>593</td>
</tr>
<tr>
<td>13</td>
<td>Wheat</td>
<td>205</td>
</tr>
<tr>
<td>14</td>
<td>Woods</td>
<td>1265</td>
</tr>
<tr>
<td>15</td>
<td>Buildings-Grass-Trees-Drives</td>
<td>386</td>
</tr>
<tr>
<td>16</td>
<td>Stone-Steel-Towers</td>
<td>93</td>
</tr>
</tbody>
</table>

4.2 Algorithm Evaluation

The cross-validation method is applied to select the procedure parameters. The polynomial kernel and Gaussian kernel are chosen as the basic kernels. In different application systems, we can choose the type of kernel functions of the offline training system. We implement the optimized multi-kernel learning on Indian Pines dataset and AVIRIS Salinas Dataset, compared with Support Vector Classifier (SVC), Kernel Sparse Representation Classifier (KSRC), SVM [44], RMKL-SVM [45] and POL-KSRC [46]. The classification results denote that the feasibility in the intelligent hyperspectral imaging instrument.

Firstly, we implement some experiments to evaluate the proposed multiple quasi-conformal kernel based on Support Vector Classifier (SVC), Kernel Sparse Representation Classifier (KSRC) on data classification. We test the single kernel and quasi-conformal multikernels for kernel classifiers on SVC and KSRC, that is, PK-SVC: Polynomial Kernel-SVC, GK-SVC: Gaussian Kernel-SVC, QMK-SVC: Quasi-conformal Multi-kernels Based SVC, PK-KSRC: Polynomial Kernel- KSRC, GK-KSRC: Gaussian Kernel- KSRC, QMK-KSRC: Quasi-conformal Multi-kernels Based KSRC. The experimental results are shown in Figure 4. Compared with the truth, QMK-SVC performs better than GK-SVC, and QMK-KSRC performs better than GK-KSRC. So the proposed method is more fitted to the hyperspectral sensing data analysis in the intelligent hyperspectral imaging instrument. In the practical applications, the multiple bands of hyperspectral data are changed to the single band of image. Thus, for the intelligent hyperspectral instrument the multiple bands of sensing data are changed to the single data.

Table 2. Performance of SVC on the Indian Pines data (%)

<table>
<thead>
<tr>
<th>Class</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>PK-SVC</td>
<td>49.3</td>
<td>58.7</td>
<td>96.4</td>
<td>39.2</td>
<td>65.8</td>
<td>93.6</td>
</tr>
<tr>
<td>GK-SVC</td>
<td>78.0</td>
<td>73.6</td>
<td>99.1</td>
<td>76.9</td>
<td>80.5</td>
<td>97.1</td>
</tr>
<tr>
<td>Class</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>PK-SVC</td>
<td>62.9</td>
<td>85.3</td>
<td>100</td>
<td>65.8</td>
<td>72.3</td>
<td>58.4</td>
</tr>
<tr>
<td>GK-SVC</td>
<td>79.7</td>
<td>89.8</td>
<td>99.7</td>
<td>83.6</td>
<td>86.0</td>
<td>80.7</td>
</tr>
</tbody>
</table>

Table 3. Performance of KSRC on the Indian Pines data (%)

<table>
<thead>
<tr>
<th>Class</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>PK-KSRC</td>
<td>51.8</td>
<td>59.6</td>
<td>96.1</td>
<td>49.1</td>
<td>78.5</td>
<td>93.8</td>
</tr>
<tr>
<td>GK-KSRC</td>
<td>77.8</td>
<td>76.4</td>
<td>99.1</td>
<td>75.5</td>
<td>79.0</td>
<td>97.4</td>
</tr>
<tr>
<td>Class</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>PK-KSRC</td>
<td>62.8</td>
<td>84.7</td>
<td>100</td>
<td>67.5</td>
<td>75.2</td>
<td>60.7</td>
</tr>
<tr>
<td>GK-KSRC</td>
<td>82.7</td>
<td>88.7</td>
<td>100</td>
<td>83.9</td>
<td>86.3</td>
<td>81.1</td>
</tr>
</tbody>
</table>
Table 4. Performance on the Pavia University data (%)

<table>
<thead>
<tr>
<th>Class</th>
<th>POL-SVM</th>
<th>RBF-SVM</th>
<th>SMKL-SVM</th>
<th>RMKL-SVM</th>
<th>POL-KSRC</th>
<th>RBF-KSRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>83.97</td>
<td>84.46</td>
<td>84.25</td>
<td>88.24</td>
<td>84.94</td>
<td>85.43</td>
</tr>
<tr>
<td>p2</td>
<td>85.25</td>
<td>91.42</td>
<td>90.26</td>
<td>86.23</td>
<td>92.46</td>
<td></td>
</tr>
<tr>
<td>p3</td>
<td>70.24</td>
<td>74.15</td>
<td>74.67</td>
<td>71.25</td>
<td>75.17</td>
<td></td>
</tr>
<tr>
<td>p4</td>
<td>87.25</td>
<td>97.67</td>
<td>97.25</td>
<td>97.27</td>
<td>98.67</td>
<td></td>
</tr>
<tr>
<td>p5</td>
<td>70.45</td>
<td>78.78</td>
<td>77.48</td>
<td>71.46</td>
<td>79.71</td>
<td></td>
</tr>
<tr>
<td>p6</td>
<td>69.35</td>
<td>71.15</td>
<td>70.16</td>
<td>69.91</td>
<td>72.17</td>
<td></td>
</tr>
<tr>
<td>p7</td>
<td>76.24</td>
<td>81.53</td>
<td>80.25</td>
<td>77.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p8</td>
<td>98.57</td>
<td>99.46</td>
<td>99.26</td>
<td>98.86</td>
<td>99.98</td>
<td></td>
</tr>
</tbody>
</table>

The experimental results on hyperspectral image databases show that the optimized multiple kernels-based machine learning achieves better performance than those of other methods on the hyperspectral data analysis for the intelligent hyperspectral imaging instrument. The kernel-based learning machine is to solve the data mapping through the selection of function and parameters of kernels. The optimized multiple kernels are combined to more precisely characterize the hyperspectral sensing data from the intelligent hyperspectral instrument for improving the performance of solving complex visual learning tasks. The experimental results show that the proposed framework outperforms others on hyperspectral data analysis for the intelligent hyperspectral instrument. In the practical application the multiple bands of hyperspectral data are changed to the single band of image. Thus, for the intelligent hyperspectral instrument the multiple bands of sensing data are changed to the single data. The classification results denote that the feasibility in the intelligent hyperspectral imaging instrument. The performance of the practical application is evaluated with the classification result. In the system, the cross-validation method is applied to select the procedure parameters, and the polynomial kernel and Gaussian kernel are chosen as the basic kernels. The type of kernel function in the offline training system is chosen through cross-validation method.

4.3 Performance Evaluation

In the part of deep learning, the iteration times is 2000; the nodes of the tree layer are 500, 500, 2000. In the SVM part we use the LIBSVM to solve this optimization problem and the parameters applied are default values. Due to the amount of the data and the effectiveness of the method, the ratio between training samples and test samples within the class is set up as 10:34.

We have conducted experiments on Indian Pines using deep learning solely with iteration number of 1, 10, 200, 2000, 3000, 4000. The Accuracy have reached the top when the iteration number is 2000 which is 45.25%. Considering that deep learning’s goal is to classify all the materials at the same time, it is understandable that the accuracy is hard to be improved. However, our goal is to recognize different kinds of materials from hyperspectral images, it is effectively unnecessary to classify them all at the same time. In other words, single classification is capable of satisfying our requirement.

The recognition accuracy result of 14 classes of materials using deep learning and single-class SVM are presented in Figure 5.

Figure 5. Recognition accuracy of single-class SVM

Some conclusions can be obtained from Figure 6, First, DLSVM perform better result than those reported previously: simple Euclidean and LOOC-based method [47-48]. The average accuracy have reached 83.99%. Comparing to the accuracy of deep learning, DLSVM have accomplished a great improvement.

Figure 6. Recognition accuracy of different classifiers

4.4 Discussion

Judging from Figure 5, the performance of our method varies upon different materials. However, most
Hyperspectral images. On multiclass SVM formulations and other substitution from softmax to SVM appears to be better than softmax on standard dataset Indian Pines. 

5 Conclusion

In conclusion, we have shown that DLSVM works better than softmax on standard dataset Indian Pines and the result exhibits the efficiency of DLSVM upon recognition of hyperspectral images. Besides, the substitution from softmax to SVM appears to be smooth and steady. Further research could be focused on multiclass SVM formulations and other hyperspectral images.

References


**Biographies**

Yifan Li is doctoral candidate at Harbin Institute of Technology Department: Innovative Information Industry Research Center of Computer Science and technology. Research Field: Deep learning, Kernel Learning, Hyperspectral image processing.
Junbao Li is professor and Ph.D. supervisor at Harbin Institute of Technology. Department: Electronics and Information Engineering. Research Field: Image processing and Pattern Recognition, Artificial Intelligent.
