Adaptive Classification of Hyperspectral Image

Hamid Dehghani & Hassan Ghassemian
Tarbiat Modaress University
P. O. Box: 14115-111, Tehran, Iran.
ghassemi@modares.ac.ir

Abstract:
An important problem in pattern recognition is the effect of limited training samples on classification performance. When the ratio of the number of training samples to the dimensionality is small, parameter estimates become highly variable, causing the deterioration of classification performance. This problem has become more prevalent in remote sensing with the emergence of a new generation of sensors. While the new sensor technology provides higher spectral and spatial resolution, enabling a greater number of spectrally separable classes to be identified, the needed labeled samples for designing the classifier remain difficult and expensive to acquire. In this paper, we propose an adaptive classification model that operates based on decision fusion. This method uses soft learning strategy. In this classifier, learning is performed at two steps. At the beginning of this method, observation space is parted and several groups of bands are produced. After providing the primary decisions, several rules are used in decision fusion center to determine the final class of pixels. Reported results on remote sensing images show classification performance is improved, and this method may solve the limitation of training samples in the high dimensional data and the Hughes phenomenon may be mitigated.

Keywords: Adaptive Classifier, Hyperspectral Data, Limited Training Samples.

1. Introduction

Remote sensing technology involves the measurement and analysis of the electromagnetic radiation reflected from the earth's surface by a passive or an active source. The radiation responses in various wavelengths indicate the types or properties of the materials on the surface being measured and collectively form a hyperspectral image. Classification is one of the most often used quantitative data analysis techniques to describe ground cover types or material classes in remote sensing data. Recent advances in imagery sensor technology have increased the spectral resolution of remote sensing data significantly. The availability of a large number of spectral bands should allow more detailed classes to be identified with higher accuracy. However, a large number of features produce large variance in estimated parameters. Typically, the performance of classifier improves up to a certain point as additional features are added, and then deteriorates. This is referred to as Hughes phenomenon [1]. In this work, we propose a new classification method for improving the influence of limitation of training samples in the hyperspectral data and the Hughes phenomenon.

A typical supervised classification system for hyperspectral data consists of several stages as shown in Figure 1. In supervised classification methods is assumed that the parameters of classifier can be estimated by the training samples. For accurate estimation and therefore accurate and reliable classification, enough training samples are necessary. For each class, in some classifiers such as MLC, the number of image bands determines the necessary minimum number of training samples. Since the number of image bands in the mentioned cases is large, even this minimum margin would be large and providing that, would be hard and expensive. Therefore, the performance of traditional classification methods isn’t useful. The effect of limited training samples could be seen in Figure 1. This problem is sharp, if we have to use multisatellite sensors and multitemporal images [2].
In this paper an adaptive classifier for mitigating the small training sample problem, is proposed. This adaptive classifier that is based on decision fusion, enhances estimation and hence improves classification accuracy by utilizing the classified samples (referred as semilabeled samples), in addition to the original training samples. This proposed adaptive classifier potentially has several benefits. First, this approach can start with a small number of training samples. Second the large number of semilabeled samples can enhance the estimation of the parameters, decreasing the estimation error and therefore reduce the effect of the small sample size problem. Third the estimated parameters are more representative of the true class distribution, because samples used to estimate these are from a larger portion of the entire data set. Fourth this classifier is adaptive and can be improved when the new data of the considered scene is available [3].

Decision fusion can be defined as the process of fusing information from individual data sources after each data source has undergone primary classifications [2]. We have to utilize some preprocessing such as, radiometric and geometric correction and registration, before using the decision fusion methods. In this paper we suppose that these preprocessing are done and provided data from different sensors are similar and have same scale. Also maximum likelihood and back-propagation neural network classifier are used as primary classifiers. After this step, in addition to proposed model, several neural and statistical decision fusion schemes are applied to make the final decision for the class of pixels.

The paper is organized into five sections. The next section provides a description about classification based on decision fusion. In Section 3, adaptive proposed classifier is described. The data sets used in the experiments and the results obtained are reported in Section 4. Finally, after discussion, conclusions are drawn in Section 5.

2. Use of Decision Fusion for Classification

In general, decision fusion methods are used for combining the local decisions of different sensors. It has been demonstrated in many applications that the use of decision fusion algorithms to combine multiple experts (individual classifiers) can enhance the classification process performance. The reason for this enhancement lies in the way decision fusion approaches can exploit additional information extracted from the data sets.

Step-1: To classify in this method, under mentioned flow chart is following. The existing data in each source is used for primary classification. The maximum likelihood classifier (MLC) and neural network classifier (NNC) are used as primary classification in this paper. MLC is the most common supervise classification. In this classifier it is assumed that the probability distributions for the classes are of the form of multivariate normal models.

If the distribution functions of the information class are unknown, nonparametric classifiers such as neural networks have better performance. Three-layer back-propagation network is used for this purpose. A squared-error cost function is defined for this classifier [4].

$$\Delta = E \left\{ \sum_{i=1}^{M} \left| y_i(x) - d_i \right|^2 \right\}$$  \hspace{1cm} (1)

Where \(E\{\cdot\}\)is the expectation operator, \(M\) is the number of classes and, \(y_i\) and \(d_i\) are real and desired output respectively, for \(i\)'th node in the neural network. For minimization this cost function, it is proved that the conditional expectation are the following.

$$E[d_i | x] = p(w_i | x)$$  \hspace{1cm} (2)

In other words, for each pixel, the neural networks can estimate the posterior probability of each class.

Step-2: For each source the output of these classifiers would be posterior probability, which specifies the degree of dependency of pixels to the classes of the given source. These probabilities will
be used for determining the class of pixels in a decision fusion center. In decision fusion center, a rule is used to make the final decision about the class of pixel.

2.3 Decision Fusion Center Rule

Rule 1: Arithmetic Mean: In this rule, the arithmetic mean of the posterior probabilities related to each class is calculated [5].

\[ C_j(X) = \sum_{i=1}^{N} p(w_i/x_j) \]  

(3)

The class, for which \( C \) is largest, is selected as the class of pixel.

Rule 2: geometric mean: The geometric mean of the posterior probabilities related to each class is calculated [6]. The considered pixel is assigned to the class that it’s \( F \) is largest.

\[ F_j(X) = \prod_{i=1}^{N} p(w_i/x_j) \]  

(4)

Rule 3: Max: Pixel \( x \) Assign to class \( w_j \), when the below equation would be true [7], [8].

\[ \max_{i=1}^{N} p(w_j/x_i) = \max_{k=1}^{M} \max_{i=1}^{N} \left[ p(w_k/x_i) \right] \]  

(5)

Where, \( M \) is the number of class in image. This method is put in practice a fuzzy model. In this fuzzy model the premises of rules are aggregated together with OR operator.

Rule 4: Neural Network: The upper rules use specific law for combining the primary decisions and final deciding. But the Neural network method is nonparametric rule. Three-layer back-propagation network is used for this purpose. The inputs of this network are the posterior probabilities provided by primary classifiers. After training, this network can determine the class of pixel [5].

3. Proposed Method

As mentioned in previous section, decision fusion methods are used in multisensor systems. But we pursue an approach that can be used for hyperspectral data classification as well as multisensor data classification. For this purpose, we consider all bands of image, careless their sensors. Hence we will deal with the classification problem of high dimensional data. With this strategy, even we can use the decision fusion rules for classifying the data of one sensor. To classify in this method, we follow the under mentioned flow chart.

Step.1: If the given scene were observed in \( L \) bands, the bands would be categorized according to different criterions such as minimum and maximum correlation. The number of bands in each group depends on total number of training samples.

Step.2: The bands of each group are considered as the bands of a new source. The existing data in each source is used for primary classification. In this step, only existing training samples are used. In this paper two simultaneous three-layer back-propagation NNC and MLC are used. Where \( M \) is the number of classes, \( K \) is the number of source and \( P \) is the number of primary classifiers. If after this step we want to use of fusion center, the equations (3)-(5) change as follow:

\[ C_j(X) = \sum_{k=1}^{K} \sum_{i=1}^{P} p_i(w_j/x_k) \]  

(6)

\[ L_j(X) = \prod_{k=1}^{K} \prod_{i=1}^{P} p_i(w_j/x_k) \]  

(7)

\[ \max_{l=1}^{K} \max_{i=1}^{P} p_i(w_j/x_l) = \max_{k=1}^{M} \max_{l=1}^{K} \max_{i=1}^{P} p_i(w_k/x_l) \]  

(8)

Step.3. After primary classification, the class of all pixels is determined in each source.

Step.4. The image is observed thoroughly and the pixels, which absolute majority primary classifiers have agreed on their class, are determined and marked.

Step.5. The number of marked samples in each class is determined and the class that having the minimum number of marked samples is chosen.

Step.6. New training or semilabeled samples for classes are selected from the marked pixels, the number of which would be the same as the minimum number specified in the previous step.

Step.7. In the decision fusion center we use a three-layer back-propagation neural network to make the final decision for the class of pixels. To train this network we use new training samples in addition to original training samples.

The functionality block diagram of this classifier is shown in Figure 2.

4. Experiments

For comparison the different methods, an experiment was developed. The multispectral data used in this experiment, is an agricultural segment of Indiana State. This image has provided in 12 bands and its radiometric resolution is 8 bit or 256 gray levels. These bands are presented in table 1. Training and testing
regions have selected from this image in 8 classes and shown as in Figure 3(a). In this test image 18 pixels per class used as training samples and shown with brown color in Figure 3(a).

For having a comparison of the performance and results of the different classifiers, we have to use some measures. Accuracy and reliability are the important measures that are applied in this paper. Accuracy is defined as:

\[ \alpha = \frac{a}{A} \]  

(9)

Where, “a” is the number of test samples that are correctly classified and A is the total test pixels. Reliability can be expressed as:

\[ \beta = \frac{a}{B} \]  

(10)

Where, B is the total test pixels.

In Conventional Methods, all bands were used together, and classification was performed with MLC and three-layer back-propagation neural network classifiers. In neural network classifier for facilitating the training process, we used gray code instead of gray level. Hence the total number of input neurons was 12*8=96. Since the number of information classes was eight, the number of output neurons was either selected as 8, and network with 16 hidden units was completed. The result of these classifiers have presented in Figure 2.

In decision fusion methods, initially it was necessary that the observed bands were categorized according to minimum and maximum correlation. The results of this process were three new sources that every one has 4 bands. The sources and their bands have presented in table 2. For primary classification in each source, both neural network and maximum likelihood classifiers were used, simultaneously. The used neural network in this step was three-layer back-propagation network. For facilitating the training process, gray code was used. Hence the number of input neurons was 4*8=32. The number of output units was 8, and the number of hidden neurons was 16.

In adaptive model and rule 5 and proposed method, we used a three-layer back-propagation network in decision fusion center. Because we have three sources and each source was classified with two methods, therefore the number of input neurons was 3*2*8=48. The number of output neurons was 8, and for hidden layer, 36 neurons were selected. The results of classification methods have presented in Figure 4 and the class maps produced by some methods have shown in Figure 3.

5. Conclusions

As the number of spectral bands of high spectral resolution data increases, the capability to detect more detailed classes should also increase, and the classification accuracy should increase as well. Often the number of training samples used for supervised classification techniques is limited, thus limiting the precision with which class characteristics can be estimated. As the number of spectral bands becomes large, the limitation on performance imposed by the limited number of training samples can become severe. In this paper, decision fusion techniques have been used to develop a supervise classification scheme for high dimensional data analysis.

The comparison of results of conventional methods in Figure 2 shows the flexibility and generalization power of nonparametric techniques, such as neural networks. Neural network approaches offer the robustness to noise and errors on training data. No general rules exist to define the network topology and to establish the procedure of the training process; moreover, it is difficult to interpret the network behavior.

In Figure 4, the results on minimum and maximum correlation criterions for categorizing the bands in image report that the minimum correlation is the better criterion for creation the new sources. In maximum correlation state, the information of these bands has maximum overlapping in comparison with minimum correlation state.

<table>
<thead>
<tr>
<th>Band</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spectral bound ((\mu))</td>
<td>0.46</td>
<td>0.48</td>
<td>0.50</td>
<td>0.52</td>
<td>0.54</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>0.49</td>
<td>0.51</td>
<td>0.54</td>
<td>0.57</td>
<td>0.60</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>0.61</td>
<td>0.72</td>
<td>1.00</td>
<td>1.50</td>
<td>2.00</td>
<td>9.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>0.92</td>
<td>1.40</td>
<td>1.80</td>
<td>2.60</td>
<td>11.7</td>
</tr>
</tbody>
</table>

Table 1. The band of image.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Source-1</th>
<th>Source-2</th>
<th>Source-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bands</td>
<td>1,2,3,7</td>
<td>4,5,6,12</td>
<td>8,9,10,11</td>
</tr>
<tr>
<td>Bands</td>
<td>1,6,8,10</td>
<td>2,4,11,12</td>
<td>3,5,7,9</td>
</tr>
</tbody>
</table>

Table 2. The new sources and their bands after grouping

<table>
<thead>
<tr>
<th>Figure 2. The results of experimentation for conventional methods</th>
<th>Figure 3. The class maps produced by some methods</th>
</tr>
</thead>
</table>
Therefore the results of primary classification of minimum correlation sources are more accurate and hence in decision fusion center the final decisions about classes of pixels are made more accurate. The methods and rules in decision fusion center will be desired that aren’t sensitive to the band categorizing criterions. The comparison of the results in Figure 4, show the effectiveness of the adaptive model. This improvement occurred, because for training the network in decision fusion center, the large number of semilabeled samples was used and on the other hand, these samples are from a larger portion of the entire data set, hence the weights of this network are more assured. In other words, the extracted information by primary classifiers can help to create the more accurate decision fusion center. Certainly we can use the marked samples to complete the training the primary classifiers. Probably, after doing this work, the provided results are more accurate. Of course we must care to time of process and error propagation. In this case, if mislabeled pixels are belonged to semilabeled samples error strongly is propagated in all parts of model with respect to experimented case.

References


