Hyperspectral Remote Sensing Image Classification Based on Rotation Forest

Junshi Xia, Student Member, IEEE, Peijun Du, Senior Member, IEEE, Xiyan He, and Jocelyn Chanussot, Fellow, IEEE

Abstract—In this letter, an ensemble learning approach, Rotation Forest, has been applied to hyperspectral remote sensing image classification for the first time. The framework of Rotation Forest is to project the original data into a new feature space using transformation methods for each base classifier (decision tree), then the base classifier can train in different new spaces for the purpose of encouraging both individual accuracy and diversity within the ensemble simultaneously. Principal component analysis (PCA), maximum noise fraction, independent component analysis, and local Fisher discriminant analysis are introduced as feature transformation algorithms in the original Rotation Forest. The performance of Rotation Forest was evaluated based on several criteria: different data sets, sensitivity to the number of training samples, ensemble size and the number of features in a subset. Experimental results revealed that Rotation Forest, especially with PCA transformation, could produce more accurate results than bagging, AdaBoost, and Random Forest. They indicate that Rotation Forests are promising approaches for generating classifier ensemble of hyperspectral remote sensing.

Index Terms—Classification, decision tree, ensemble learning, hyperspectral remote sensing image, Rotation Forest.

I. INTRODUCTION

HYPERSPECTRAL remote sensing image classification is a challenging problem because of its high dimensional inputs (hundreds of bands), many class outputs, and limited availability of reference data [1], [2]. Therefore, we require some powerful techniques to improve the accuracy of classification results. Since it is always difficult to select an optimal classifier, an attractive type of machine learning algorithm called the multiple classifier system (MCS) or classifier ensemble is rapidly developing and enjoying a lot of attentions due to their potential to improve the classification accuracy of hyperspectral remote sensing image significantly [1], [3].

MCS integrates the outputs of individual classifiers according to a certain combination approach (such as majority vote, Bayesian rule, etc.) or based on an iterative error minimization ([3]–[7]). The outputs can be generated by the same classifier of different training samples, or the different classifiers of same/different training set. Previous studies demonstrated that a successful MCS should be one where the member classifiers are accurate as well as the diversities among them are obvious, because combining similar classification results would not further improve the accuracy [8]–[10]. Two popular approaches for creating strong ensembles are boosting and bagging [4], [11]. Boosting processes data with iterative re-training, and the weights of misclassified samples are increased to concentrate the learning algorithm on specific samples [4], [12]. In contrast, bagging can produce accurate ensemble by training many classifiers on boot-strapped samples from training set [11]. Diversity in bagging is provided with further randomization yielding Random Forest ensemble approach [13]. Random Forest adopts decision trees trained on bootstrap samples and the diversity is promoted with random choice of features at each node while constructing the trees. It can overcome the drawbacks of bagging and boosting algorithms (e.g., high computational cost and sensitivity to noise) [7]. In addition, limiting the number of variables in Random Forest used for a split, the computational complexity can be reduced and the correlation between the trees be decreased. This enables Random Forest to deal with high-dimensional datasets [9].

Rotation Forest, proposed by Rodriguez et al. [14] is based on the idea of Random Forest. The main idea of Rotation Forest is to encourage simultaneously both member diversities and individual accuracy within a classifier ensemble. In the framework of Rotation Forest, each classifier is independently constructed using decision tree method, and each tree is trained on the training samples in a rotated feature space derived from principal component analysis (PCA) transformation. One of the most important point of ensemble methods is to select the base classifier. Decision tree is always used for rotation task because of its sensitivity to rotation of the feature axes [15]. Though Rotation Forest performs much better than other ensemble methods (bagging, AdaBoost, Random Forest) on some benchmark classification from UCI repository [14], [16], the performance for classify hyperspectral remote sensing im-
The success of Rotation Forest relies on the base classifier and the rotation matrix created by the transformation methods. Decision tree is always adopted for Rotation Forest because it is sensitive to rotation of the axes. Here, we selected classification and regression tree (CART) as the base classifier [17].

CART is based on the Gini index, which is treated as node impurity criterion [9], [17]

\[ Gini(t) = \sum_{i=1}^{c} P_{\omega_i} (1 - P_{\omega_i}) \]  

where \( c \) is the number of classes and \( P_{\omega_i} \) is the probability of class \( \omega_i \) at node \( t \). \( P_{\omega_i} \) is defined as 

\[ P_{\omega_i} = \frac{n_{\omega_i}}{N} \]  

where \( n_{\omega_i} \) is the number of samples of class \( \omega_i \) and \( N \) is the total number of training samples.

The summed Gini index selects the split that maximizes the decrease in impurity. By employing this rule, CART generates a sequence of subtrees by growing a large tree and pruning it back until only the root node is left. Then it uses cross-validation to estimate the misclassification cost of each subtree and chooses the one with the lowest estimated cost [18].

In [19], the authors compared the performance of different transformation algorithms (e.g., PCA, NDA, and RP) and found that PCA produced the best results. In this letter, we will further examine the efficiency of common transformation algorithms applied to hyperspectral remote sensing image classification, such as PCA [20], maximum noise fraction (MNF) [21], [22], independent component analysis (ICA) [23], [24], and LFDA. PCA, MNF, and ICA are all unsupervised feature extraction methods, while LFDA is supervised. PCA and MNF maximize the amount of data variance and signal-to-noise ratio (SNR), respectively, yielding a transformed data set in a new uncorrelated coordinate system, while ICA transforms the data into maximally independent components [20]–[22]. However, PCA, MNF, and ICA all maximize the information contained in the first transformed components, relegating variations of less significant size to low-order components [25]. LFDA effectively combines the ideas of Fisher discriminant analysis (FDA) and locality-preserving projection (LPP) [26]. That makes LFDA can both maximize between-class separability and preserves with-class local structure. More details about LFDA can be seen in [26]. In order to preserve the variability information in the images, all components using the above three transformation methods are retained.

### III. Experiments and Result Analysis

In order to assess the performance of Rotation Forest algorithm, we conduct the experiments with three widely used hyperspectral images obtained from NASA’s Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), Reflective Optics System Spectrographic Imaging System (ROSIS), and Digital Airborne Imaging Spectrometer (DAIS) owned by the German Aerospace Center (DLR). AVIRIS dataset is captured over a vegetation area of Indian Pines, Indiana, USA.
Among them, LORSAL method is the fastest. RoF (ICA) are more efficient than RoF (LFDA) and SVM. The computational time also included the time consumed on processors, 12 GB memory. In case 2, the computation time of RoF (PCA, RoF (ICA), RoF (MNF), RoF (LFDA)) and SVM, LORSAL was 5.38 s, 3.18 s, respectively. The computation time of Rotation Forest gave the better performance than SVM in all cases. In addition, we have compared the computation time of these methods on an Intel(R) Xeon(R) CPU X5660 @ 2.80 GHz, 2.79 GHz, two processors, 12 GB memory. In case 2, the computation time of RoF (PCA), RoF (MNF), RoF (ICA), RoF (LFDA), SVM, LORSAL was 8.53 s, 9.28 s, 9.16 s, 36.84 s, 42.65 s, and 3.18 s, respectively. The computational time of Rotation Forest approaches is longer than the one of bagging, AdaBoost, and Random Forest. For the approaches of SVM and RoF (LFDA), the computational time also included the time consumed on the parameter determination. So RoF (PCA), RoF (MNF), and RoF (ICA) are more efficient than RoF (LFDA) and SVM. Among them, LORSAL method is the fastest.

Table I shows the classification accuracies (OA%) obtained by the Rotation Forest approaches as well as other algorithms using different training samples. We highlight the highest OA of each case in bold font. It can be seen that RoF (PCA, ICA) achieve good results than other ensemble approaches (bagging, AdaBoost, and Random Forest), where the OA is always increased as the number of training samples is increased. For instance, in case 1, CART, bagging, AdaBoost, and RF acquired an OA of 57.25%, 66.5%, 66.98%, and 71.38, respectively, RoF (PCA) and RoF (ICA), respectively, increased the OA to 79.65% and 76.78%, while the OA of RoF (MNF) and RoF (LFDA) were improved to 76.78% and 71.66%. LORSAL yielded the highest OA and RoF (PCA) gave the better performance than SVM in all cases. In addition, we have compared the computation time of these methods on an Intel(R) Xeon(R) CPU X5660 @ 2.80 GHz, 2.79 GHz, two processors, 12 GB memory. In case 2, the computation time of RoF (PCA), RoF (MNF), RoF (ICA), RoF (LFDA), SVM, LORSAL was 8.53 s, 9.28 s, 9.16 s, 36.84 s, 42.65 s, and 3.18 s, respectively. The computational time of Rotation Forest approaches is longer than the one of bagging, AdaBoost, and Random Forest. For the approaches of SVM and RoF (LFDA), the computational time also included the time consumed on the parameter determination. So RoF (PCA), RoF (MNF), and RoF (ICA) are more efficient than RoF (LFDA) and SVM. Among them, LORSAL method is the fastest.

A. Results of Indian Pines AVIRIS Image

Table II shows the classification accuracies (OA%) obtained by the Rotation Forest approaches as well as other algorithms using different training samples. We highlight the highest OA of each case in bold font. It can be seen that RoF (PCA, ICA) achieve good results than other ensemble approaches (bagging, AdaBoost, and Random Forest), where the OA is always increased as the number of training samples is increased. For instance, in case 1, CART, bagging, AdaBoost, and RF acquired an OA of 57.25%, 66.5%, 66.98%, and 71.38, respectively, RoF (PCA) and RoF (ICA), respectively, increased the OA to 79.65% and 76.78%, while the OA of RoF (MNF) and RoF (LFDA) were improved to 76.78% and 71.66%. LORSAL yielded the highest OA and RoF (PCA) gave the better performance than SVM in all cases. In addition, we have compared the computation time of these methods on an Intel(R) Xeon(R) CPU X5660 @ 2.80 GHz, 2.79 GHz, two processors, 12 GB memory. In case 2, the computation time of RoF (PCA), RoF (MNF), RoF (ICA), RoF (LFDA), SVM, LORSAL was 8.53 s, 9.28 s, 9.16 s, 36.84 s, 42.65 s, and 3.18 s, respectively. The computational time of Rotation Forest approaches is longer than the one of bagging, AdaBoost, and Random Forest. For the approaches of SVM and RoF (LFDA), the computational time also included the time consumed on the parameter determination. So RoF (PCA), RoF (MNF), and RoF (ICA) are more efficient than RoF (LFDA) and SVM. Among them, LORSAL method is the fastest.

Table II shows the classification accuracies (OA%) obtained by the Rotation Forest approaches as well as other algorithms using different training samples. We highlight the highest OA of each case in bold font. It can be seen that RoF (PCA, ICA) achieve good results than other ensemble approaches (bagging, AdaBoost, and Random Forest), where the OA is always increased as the number of training samples is increased. For instance, in case 1, CART, bagging, AdaBoost, and RF acquired an OA of 57.25%, 66.5%, 66.98%, and 71.38, respectively, RoF (PCA) and RoF (ICA), respectively, increased the OA to 79.65% and 76.78%, while the OA of RoF (MNF) and RoF (LFDA) were improved to 76.78% and 71.66%. LORSAL yielded the highest OA and RoF (PCA) gave the better performance than SVM in all cases. In addition, we have compared the computation time of these methods on an Intel(R) Xeon(R) CPU X5660 @ 2.80 GHz, 2.79 GHz, two processors, 12 GB memory. In case 2, the computation time of RoF (PCA), RoF (MNF), RoF (ICA), RoF (LFDA), SVM, LORSAL was 8.53 s, 9.28 s, 9.16 s, 36.84 s, 42.65 s, and 3.18 s, respectively. The computational time of Rotation Forest approaches is longer than the one of bagging, AdaBoost, and Random Forest. For the approaches of SVM and RoF (LFDA), the computational time also included the time consumed on the parameter determination. So RoF (PCA), RoF (MNF), and RoF (ICA) are more efficient than RoF (LFDA) and SVM. Among them, LORSAL method is the fastest.
TABLE III
Overall Accuracies (%) for the University of Pavia ROSIS Image Using Different Number of Training Samples

<table>
<thead>
<tr>
<th>Number of Training sample</th>
<th>CART</th>
<th>Bagging</th>
<th>Adaboost</th>
<th>RF</th>
<th>RoF (PCA)</th>
<th>RoF (ICA)</th>
<th>RoF (MNF)</th>
<th>RoF (LFDA)</th>
<th>SVM</th>
<th>LORSAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>59.53</td>
<td>67.78</td>
<td>66.83</td>
<td>66.77</td>
<td>78.38</td>
<td>71.92</td>
<td>70.45</td>
<td>73.02</td>
<td>76.42</td>
<td>71.28</td>
</tr>
<tr>
<td>Case 2</td>
<td>62.82</td>
<td>68.26</td>
<td>67.8</td>
<td>68.9</td>
<td>80.71</td>
<td>76.37</td>
<td>71.48</td>
<td>75.2</td>
<td>77.35</td>
<td>76.07</td>
</tr>
<tr>
<td>Case 3</td>
<td>63.39</td>
<td>69.64</td>
<td>70.13</td>
<td>69.9</td>
<td>82.89</td>
<td>75.91</td>
<td>72.59</td>
<td>75.73</td>
<td>77.86</td>
<td>78.08</td>
</tr>
<tr>
<td>Case 4</td>
<td>64.93</td>
<td>70.11</td>
<td>70.3</td>
<td>71.11</td>
<td>83.14</td>
<td>78.04</td>
<td>73.28</td>
<td>75.57</td>
<td>79.94</td>
<td>80.09</td>
</tr>
</tbody>
</table>

TABLE IV
Overall Accuracies (%) for the Center of Pavia DAIS Image Using Different Number of Training Samples

<table>
<thead>
<tr>
<th>Number of Training sample</th>
<th>CART</th>
<th>Bagging</th>
<th>Adaboost</th>
<th>RF</th>
<th>RoF (PCA)</th>
<th>RoF (ICA)</th>
<th>RoF (MNF)</th>
<th>RoF (LFDA)</th>
<th>SVM</th>
<th>LORSAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>87.95</td>
<td>90.89</td>
<td>91.83</td>
<td>93.12</td>
<td>95.64</td>
<td>95.2</td>
<td>95.06</td>
<td>95.52</td>
<td>93.05</td>
<td>93.76</td>
</tr>
<tr>
<td>Case 2</td>
<td>90.51</td>
<td>91.49</td>
<td>92.45</td>
<td>93.03</td>
<td>95.76</td>
<td>95.36</td>
<td>94.91</td>
<td>95.6</td>
<td>94.17</td>
<td>94.45</td>
</tr>
<tr>
<td>Case 3</td>
<td>91.25</td>
<td>92.09</td>
<td>92.22</td>
<td>93.95</td>
<td>95.78</td>
<td>95.64</td>
<td>95.15</td>
<td>95.57</td>
<td>94.67</td>
<td>94.47</td>
</tr>
<tr>
<td>Case 4</td>
<td>91.57</td>
<td>92.17</td>
<td>92.61</td>
<td>94.8</td>
<td>95.81</td>
<td>95.48</td>
<td>95.28</td>
<td>95.52</td>
<td>95.1</td>
<td>94.6</td>
</tr>
</tbody>
</table>

B. Results of University of Pavia ROSIS Image

Table III reports the OA for each approaches using different number of training samples. From Table III, it can be observed that RoF (PCA) outperform other algorithms in all cases. RoF (LFDA) gave the better performance than RoF (MNF). In case 4, the corresponding OA of RoF (PCA) achieved on the test set was 83.14%, higher than the one of SVM (79.98%) and LORSAL(80.09%). More details can be seen in Table III. The computation complexity is similar with the previous AVIRIS experiment. The computation complexity of RoF (PCA), RoF (MNF), and RoF (ICA) are much less than the one of SVM and RoF (LFDA). And LORSAL method is more effective than RoF (PCA) algorithm.

C. Results of Center of Pavia DAIS Image

The global accuracies of different method using different number of samples are reported in Table IV. The Pavia Center DAIS data set was easy to classify since even the CART acquires extraordinarily high classification accuracy. Regarding the global accuracies, Rotation Forest with different transformation algorithms are all superior to other compared approaches. In case 1–3, RoF (PCA) achieved the best global accuracies with the OA (95.38%). And RoF (LFDA) yielded the highest OA (95.02%) in case 4.

D. Sensitivity of Parameters

Ensemble size ($L$) and the number of features in a subset ($M$) are the key parameters of Rotation Forest, also known as an indicator of the operating complexity. In order to investigate the impacts of these parameters, we have performed the classification results using different ensemble size when the number of subset $M$ is fixed to 10, different number of features in a subset when ensemble size $L$ is fixed to 10. Fig. 1 shows the OA (%) using different number of $L$ and $M$ obtained from ROSIS image (case 1). With the increment of $L$, the classification results are significantly improved. Different approaches achieved the best classification result at different number of $L$ and $M$. For instance, when $M$ was fixed, RoF (PCA) obtained the best OA when $L$ equals to 10. Furthermore, we also conducted the similar experiments on AVIRIS and DAIS images.

IV. CONCLUSION

A method for generating ensemble of classifiers, Rotation Forest, was introduced into hyperspectral remote sensing image classification. It consists in splitting the feature set into $K$ subsets, running transformation algorithms separately on
each subset and then reassembling a new extracted feature set while keeping all the components. CART decision tree classifier is used as the base classifier. Different splits of the feature set lead to different rotations. Thus diverse classifiers are obtained. Thus, we target diversity and accuracy together. We have applied Rotation Forest using different transformation approaches, including PCA, MNF, ICA, and LFDA to classify hyperspectral remote sensing image and compared with bagging, AdaBoost, Random Forest, and other advance classifiers. Experimental results have shown that RoF (PCA) outperformed other methods in terms of accuracies. The key parameters of Rotation Forest are also explored in this letter. Future studies will be focused on the integration of Rotation Forest with other ensemble approaches, the selection of an optimized decision tree model, and the use of other effective feature extraction algorithms.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe from Purdue University, West Lafayette, IN, USA, for providing the publicly available AVIRIS data and Prof. Gamba from the University of Pavia, Italy, for providing ROSIS and DAIS images.

REFERENCES