

Chapter 6

Rotation-Based Ensemble Classifiers for High-Dimensional Data

Junshi Xia, Jocelyn Chanussot, Peijun Du and Xiyan He

Abstract In past 20 years, Multiple Classifier System (MCS) has shown great potential to improve the accuracy and reliability of pattern classification. In this chapter, we discuss the major issues of MCS, including MCS topology, classifier generation, and classifier combination, providing a summary of MCS applied to remote sensing image classification, especially in high-dimensional data. Furthermore, the recently rotation-based ensemble classifiers, which encourage both individual accuracy and diversity within the ensemble simultaneously, are presented to classify high-dimensional data, taking hyperspectral and multirate remote sensing images as examples. Rotation-based ensemble classifiers project the original data into a new feature space using feature extraction and subset selection methods to generate the diverse individual classifiers. Two classifiers: Decision Tree (DT) and Support Vector Machine (SVM), are selected as the base classifier. Unsupervised and supervised feature extraction methods are employed in the rotation-based ensemble classifiers. Experimental results demonstrated that rotation-based ensemble classifiers are superior to Bagging, AdaBoost and random-based ensemble classifiers.

J. Xia · J. Chanussot · X. He
GIPSA-lab, Grenoble Institute of Technology, 38400 Grenoble, France
e-mail: junshi.xia@gipsa-lab.grenoble-inp.fr

J. Chanussot
Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik, Iceland
e-mail: jocelyn.chanussot@gipsa-lab.grenoble-inp.fr

P. Du (✉)
Department of Geographical Information Science, Nanjing University, Nanjing 210093, China
e-mail: dupjrs@gmail.com

X. He
e-mail: xiyan.he@gipsa-lab.grenoble-inp.fr

6.1 Introduction

Learning from high-dimensional data has important applications in areas such as speech processing, medicine, monitoring urbanization using multisource images, mineralogy using hyperspectral images [34, 56, 62, 64, 75, 77, 93]. Despite constant improvements in computational learning algorithms, supervised classification of high-dimensional data is still a challenge largely due to the curse of dimensionality (Hughes phenomenon) [40]. This is because the training set is very limited when compared to the hundreds or thousands of dimensions in high-dimensional data [8, 54]. Big efforts on feature extraction and feature selection have been applied to the supervised classifiers [35, 59, 61, 67, 69, 83, 92]. Since each learning algorithm (feature selection/extraction, classifiers) has its own advantages and disadvantages, efficient methodologies have yet to be developed. One of the most usual ways to achieve that is Multiple Classifier System (MCS) [4, 7, 11, 20, 27, 48, 50, 74, 78, 79, 85].

MCS comes from the idea that seek advices from several persons to make the final decision, where the basic assumption is that combining the opinions will produce a decision that is better than the single opinion [48, 50, 74]. The individual classifiers (member classifiers) are constructed and their outputs are integrated according to a certain combination approach, to gain the final classification result. The outputs can be generated by the same classifier with different training sets, or by the different classifiers with same or different training set. The success of MCS not only depends on a set of appropriate classifiers, but also on the diversity within the ensemble, which referred to two conditions: accuracy and diversity [15, 47]. Accuracy requires a set of appropriate classifiers to be as accurate as possible. Diversity means the difference among the classification results. Combining similar classification results would not further improve the accuracy. Both theoretical and empirical studies demonstrated that using a good diversity measure is able to find the extent of diversity among classifier and estimate the improvement in accuracy of combining individual classifiers [50, 74]. However, Brown et al. pointed out that the diversity for classification tasks is still an ill-defined concept, and defining an appropriate diversity measure for MCS is still an open question [12].

Generally speaking, we often adapt three independent steps: topology selection, classifier generation, and classifier combination, to construct the MCS. In Sect. 6.2, we will give a review on the uses of MCS, including these steps and along with the application of remote sensing.

Rotation-based ensemble classifier is one of the current state-of-the-art ensemble classifier methods [72]. This algorithm constructs different training sets as follows: first, the feature set is divided into several disjoint sets on which the training set is projected. Second, the subtraining set is obtained from the projection results using bootstrapping technique. Third, feature extraction is used to rotate each obtained subtraining set. The components obtained from feature extraction are rearranged to form the dataset that is treated as the input of a single individual classifier. The final result is produced by combining the output of individual classifiers generated by repeating the above steps in multiple times.

In this chapter, we will apply rotation-based ensemble classifier to classify high-dimensional data. In particular, two classifiers: Decision Tree (DT) and Support Vector Machine (SVM), are selected as the base classifiers. Unsupervised and supervised feature extraction methods are employed to rotate the training set. The performances of rotation-based ensemble classifiers are evaluated by the high-dimensional remote sensing images.

The remainder of this article is organized as follows. In Sect. 6.2, we introduce the topology, classifier generation, and classifier combination approaches of MCS, summarize the advances of MCS to high-dimensional remote sensing data classification. The main idea and two implementations of rotation-based ensemble are shown in Sects. 6.3 and 6.4, respectively. Experimental results are presented in Sect. 6.5. The conclusion and perspective of this chapter are drawn in Sect. 6.6.

6.2 Multiple Classifier System

Different classifiers, such as parametric classifiers and non-parametric classifiers, have their own strengths and limitations. The famous ‘no free lunch’ theorem stated by Wolpert may be extrapolated to the point of saying that there is no single computational view that solves all problems [86]. In the remote sensing community, Giacinto et al. compared the performances of different classification approaches in various applications and found that no one could always gain the best result [32]. In order to alleviate this problem, MCS can provide the complementary information of the pattern classifiers and integrate the outputs of these pattern classifiers so as to make the best use of the advantages and bypass the disadvantages. Nowadays MCS are highlighted by review articles as a hot topic and promising trend in remote sensing image classification and change detection [4, 21].

Most of MCS approaches focus on integrating the supervised classifiers. Few works devote to combine unsupervised classification results, often called *cluster ensemble* [38, 41]. Gao et al. proposed an interesting work to combine multiple supervised and unsupervised models using graph-based consensus maximization [29]. Unsupervised models (clustering), which do not directly generate label prediction for each individual classifier, can provide useful constraints for the joint prediction of a set of related object. Thus, Gao et al. proposed to consolidate a classification solution by maximizing the consensus among both supervised predictions and unsupervised constraints based on the optimization problem on a bipartite graph [29]. Experimental results on three real applications demonstrate the benefits of the proposed method over existing alternatives. In this chapter, we focus on the combination of supervised classifiers.

The main issues of MCS design are [50, 74, 88]:

- MCS topology: How to interconnect individual classifiers.
- Classifier generation: How to generate and select valuable classifiers.
- Classifier combination: How to build a combination function which can exploit the strengths of the selected classifiers and combine them optimally.

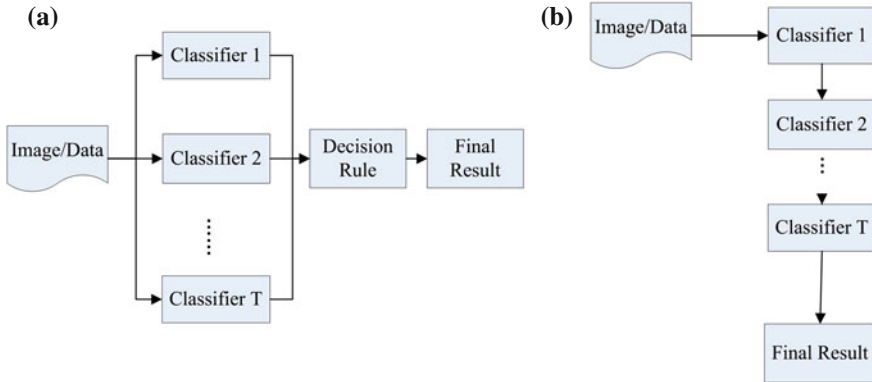


Fig. 6.1 The topologies of MCS. **a** Parallel style. **b** Concatenation style

6.2.1 MCS Topology

Figure 6.1 illustrates the two topologies employed in MCS design. The overwhelming majority of MCS reported in the literature is structured in a parallel style. In this architecture, multiple classifiers are designed independently without any mutual interaction and their outputs are combined according to certain strategies [70, 71, 90]. Alternatively, in the concatenation topology, the classification result generated by a classifier is used as the input into the next classifier [70, 71, 90]. When the primary classifier cannot obtain the satisfactory classification result, then the output of the primary classifier is feed to a secondary classifier, and so on. The main drawback of this topology is that the mistakes produced by the earlier classifier cannot be corrected by the later classifiers.

A very special case of concatenation topology is the AdaBoost [28]. The goal of AdaBoost is to enhance the accuracy of any given learning algorithm, even weak learning algorithms with an accuracy slightly better than chance. The algorithm processes training of the weak learner multiple times, each time presenting it with an updated weight over the training samples. Then, the weights of misclassified samples are increased to concentrate the learning algorithm on specific samples. Finally, the decisions generated by the weak learners are combined into a single decision.

6.2.2 Classifier Generation

Classifier generation aims to build mutually complementary individual classifiers that are accurate and at the same time disagree on some different parts of the input space. Diversity of individual classifiers is a vital requirement for the success of the MCS.

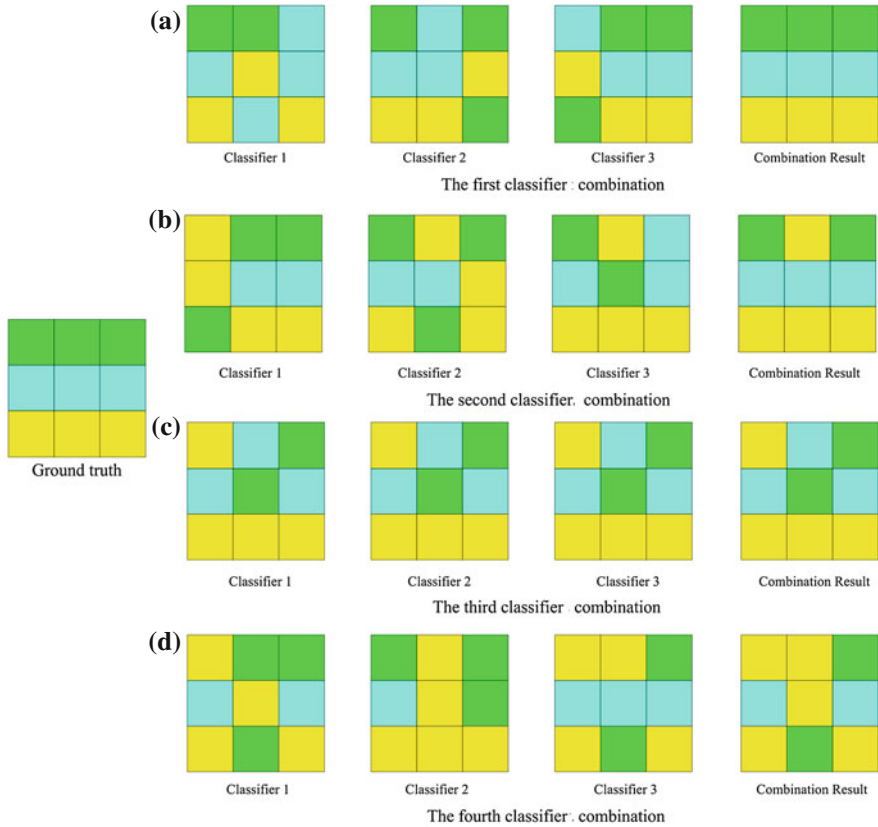


Fig. 6.2 Different classifier combinations using three single classifiers. The three colors represent the different classes. The overall accuracy of all individual classifier is 6/9. The overall accuracies of the four combinations are 1, 8/9, 6/9, and 5/9, respectively

Both theoretical and empirical studies indicate that we can ensure diversity using *Homogeneous* and *Heterogeneous* approaches [50, 74]. In *Homogeneous* approaches, we can obtain a set of classification results obtained by the same classifier by injecting randomness into the classifier, manipulating the training sample and the input features. The *Heterogeneous* approaches are to apply different learning algorithms to the same training set. First of all, we will start to review some diversity measures, and the generated classifiers followed to ensure the diversity in the ensemble.

6.2.2.1 Diversity Measures

Diversity represents the difference among the individual classifiers [15, 47]. Figure 6.2 presents four different classifier combinations within three classes (9 samples) using majority vote approach. Overall accuracy of each individual classifier is

Table 6.1 Summary of the 15 diversity measures

Name	p/n	s/dis/c	↑ / ↓	Range	Reference
Kappa statistic (κ_1, κ_2)	p	s	↓	[-1, 1]	[57, 60]
Mutual Information (MI)	p	s/c	↓	/	[43]
Q-statistic (Q)	p	s/c	↓	[-1, 1]	[49, 53]
Correlation coefficient (ρ)	p	s/c	↓	[-1, 1]	[53]
Double fault (DF)	p	s	↓	[0, 1]	[30]
Disagreement (Dis)	p	dis	↑	[0, 1]	[39]
Same fault (SF)	p	s	↓	[0, 1]	[2]
Weighted count of errors and correct (WCEC)	p	s	↑	/	[2]
Entropy (E)	n	s	↓	[0, 1]	[17, 53]
Kohavi-Wolpert variance (KW)	n	dis	↑	[0, 0.5]	[45]
Interrater agreement (IA)	n	s	↓	[0, 1]	[25, 53]
Generalized diversity (GD)	n	dis	↑	[0, 1]	[53, 68]
Conincident failure diversity (CFD)	n	dis	↑	[0, 1]	[68]
Difficulty (θ)	n	dis	↓	[0, 0.25]	[37]

Note ‘p’ stands for ‘pairwise’ and ‘n’ stands for ‘non pairwise’, ‘s’ means ‘similarity’, ‘c’ means ‘correlation’ and ‘dis’ means ‘dissimilarity.’ The arrow specifies the greater diversity if the measure is lower (↓) or higher (↑)

6/9. The accuracies of the four combinations are 1, 8/9, 6/9, and 5/9, respectively. Our goal is to use diversity measures to find the classifier combination like in Fig. 6.2a or b and avoid to select the third or especially the fourth classifier combination.

Kuncheva and Whitaker summarized the diversity measures in classifier ensembles [53]. A special issue called “Diversity Measure in Multiple Classifier System” published in Information Fusion journal indicates that diversity measure is an important research direction in MCS [51]. Petrakos et al. applied agreement measure in decision fusion level combination [60]. Foody compared the different classification results from three aspects: similarity, non-inferiority and difference using hypothesis tests and confidence interval algorithms [26]. It is proved that increasing diversity should lead to better accuracy, but there is no formal proof of this dependency [12]. Table 6.1 summarizes the 15 diversity measures with their types, data range and literature sources.

Diversity measures also play an important role in ensemble pruning. Ensemble pruning aims at reducing the ensemble size prior to combination while maintaining a high diversity among the remaining members in order to reduce the computational cost and memory storage. To deal with the ensemble pruning process, several approaches have been proposed such as clustering-based, ranking-based, and optimization-based approaches [82].

6.2.2.2 Ensuring Diversity

Following the steps of pattern classification, we can enforce the diversity by the manipulation of training samples, features, outputs and classifiers.

Manipulating the training samples: In this method, each classifier is trained on different versions of training samples by exchanging the distribution of original training samples. This method is very useful for the unstable learner (decision tree and neural network), for which small changes in the training set will lead to a major change in the obtained classifier. Bagging and Boosting belong to this category [9, 28]. Bagging applies sampling with replacement to obtain the independent training samples for individual classifiers. Boosting changed the weights of training samples according to the results of the previous trained classifiers, focusing on the wrong classified samples, making the final result using a weight vote rule.

Manipulating the training features: The most well-known algorithm of this type is Random subspace [39]. Random subspace can be employed for several types of base learners, such as DT (Random Forest) [10], SVM [85]. Another development is Attribute Bagging, which establishes the appropriate size of a feature subsets, and then creates random projections of a given training set by random selection of feature subsets [13].

Manipulating the outputs: Multiclassification problem can be converted into several two-class classification problems. Each problem discover the discrimination between one class and the other classes. Error Correcting Output Coding (ECOC) adapts a code matrix to convert a multiclass problem into binary ones. Ensemble of multiclassifier classification problem can be treated as ensembles of multiple two-classifier classification problem, and then combined together [19]. The other method to deal with the outputs is label switching [58]. This method generates an ensemble by using perturbed version of the training set where the classes of the training samples are randomly switched. High accuracy can be achieved with fairly large ensembles generated by class switching.

Manipulating the individual classifiers: We can use different classifiers or the same classifier with different parameters to ensure the diversity. For instance, when the SVM is selected as the base learner, we can gain diversity by using different kernel functions or parameters.

6.2.3 Classifier Combination

Majority vote is a simple and an effective strategy for classifier combination. Within this scheme, a pixel is assigned as the class which gets the highest vote from the individual classifiers. Foody et al. used majority vote rule to integrate multiple binary classifiers for the mapping of a specific class [27]. According to the output of individual classifier, classifier combination approaches can be divided into three levels: abstract level, rank level, and measurement level [76]. The abstract level combination methods are applied when each classifiers outputs a unique label [76]. Rank level makes use of a ranked list of classes where ranking is based on decreasing likelihood. In the measurement level, probability values of the classes provided by each classifier are used in the combination. Majority/weighted vote, fuzzy integral, evidence theory, and dynamic classifier selection belong to the abstract level com-

Table 6.2 Summary of classifier combination approaches

Name	Hard labels	Soft labels	Validation set	Reference
Majority vote	Y	N	N	[50]
Weighted vote	Y	N	Y	[63, 90]
Bayesian average	N	Y	N	[30]
Dempster-shafer evidence theory	Y	N	Y	[50, 81]
Fuzzy integral	Y	N	Y	[46, 65]
Consensus theory	Y	Y	Y	[5, 6]
Dynamic classifier selection	Y	N	Y	[31, 78, 87]

Note “Y” and “N” mean whether or not the hard labels, soft labels or validation set are needed. Dynamic classifier selection method needs the original image to calculate the distance

combination methods. Bayesian average and Consensus theory belong to measurement level methods. Table 6.2 summarizes classifier combinational approaches. Weighted vote, fuzzy integral, Dempster-Shafer evidence theory and consensus theory require another training set to calculate the weights. Dynamic classifier selection calculates the distance between the samples so it requires the original image. The computation time of dynamic classifier selection is more expensive than other approaches.

6.2.4 Applications to High-Dimensional Remote Sensing Data

Table 6.3 lists the studies of MCS applied to high-dimensional remote sensing images in recent years. These studies applied different effective MCS schemes to classify high-dimensional data, including multisource, multivariate, and hyperspectral remote sensing data. In the works of Smits [78], Briem et al. [11], Gislason et al. [33], dynamic classifier selection, Bagging, Boosting and Random Forest are applied to classify multisource remote sensing data, respectively. Lawrence et al. [55], Kawaguchi and Nishii [44], Chan and Paelinckx [14], Rodriguez-Galiano et al. [73] used Boosting and Random Forest for the classification of multi-date remote sensing images. Doan and Foody [20] combining the soft classification results derived from NOAA AVHRR images using average operator and Evidence theory. From Table 6.2, the most well-known MCS approaches for hyperspectral image classification is Random Forest. In Random Forest, each tree is trained on a bootstrapped sample of the original dataset and only a randomly chosen subset of the dimensions is considered for splitting a leaf. Thus, the computational complexity can be reduced and the correction between the trees are decreased. Apart from this, Waske et al. [85] developed random selection-based SVM for the classification of hyperspectral images. Yang et al. [91] proposed a novel subspace selection mechanism, dynamic subspace method, to improve random subspace method on automatically determining dimensionality and selecting component dimensions for diverse subspace. Du et al. [22] constructed diverse classifiers using different feature extraction methods and then combined the results using evidence theory, linear consensus algorithms. Recently,

Table 6.3 Studies on high-dimensional remote sensing image classification using MCS published in journals in recent years

Study	Methods	Datasets
Smits [78]	Dynamic classifier selection	Multispectral and SAR images
Briem et al. [11]	Bagging, Boosting and Consensus theory	Landsat MSS/AMSS+SAR and elevation, slope, aspect data
Lawrence et al. [55]	Stochastic gradient boosting	Multi-temporal Landsat TM images
Ham et al. [36]	Random Forest	Hyperspectral images
Gislason et al. [33]	Random Forest	Landsat MSS and elevation, slope, aspect data
Doan and Foody [20]	Average operator and Evidence theory	NOAA AVHRR images
Kawaguchi and Nishii [44]	AdaBoost with stump functions	Hyperspectral images
Chan and Paelinckx [14]	Random Forest and AdaBoost tree-based ensemble	Hyperspectral images
Waske et al. [84]	Random Forest	Hyperspectral images
Yang et al. [91]	Dynamic random subspace	Hyperspectral images
Waske et al. [85]	Random subspace	Hyperspectral images
Bakos and Gamba [3]	Hierarchical hybrid decision tree	Hyperspectral images
Du et al. [22]	Evidence theory, linear consensus	Hyperspectral images
Rodriguez-Galiano et al. [73]	Random Forest	Multi-temporal Landsat TM images
Xia et al. [89]	Rotation Forest	Hyperspectral images

Xia et al. [89] used Rotation Forest to classify hyperspectral remote sensing images. Compared to Random Forest, Rotation Forest [89] uses feature extraction to promote both the diversity and the accuracy of individual classifiers. Therefore, Rotation Forest can generate more accurate result than Random Forest.

6.3 Rotation-Based Ensemble Classifiers

In this study, rotation-based ensemble classifiers are used for high dimensional data. Let $\{\mathbf{X}, Y\} = \{\mathbf{x}_i, y_i\}_{i=1}^n$ be training samples. T is number of classifier. K is number of subsets (M : number of features in each subset). Γ is the base classifier. The details of rotation-based ensemble are presented in Algorithm 1 and Fig. 6.3 [66, 72]. According to Algorithm 1 and Fig. 6.3, the main steps of rotation-based ensemble classifier can be concluded as follows:

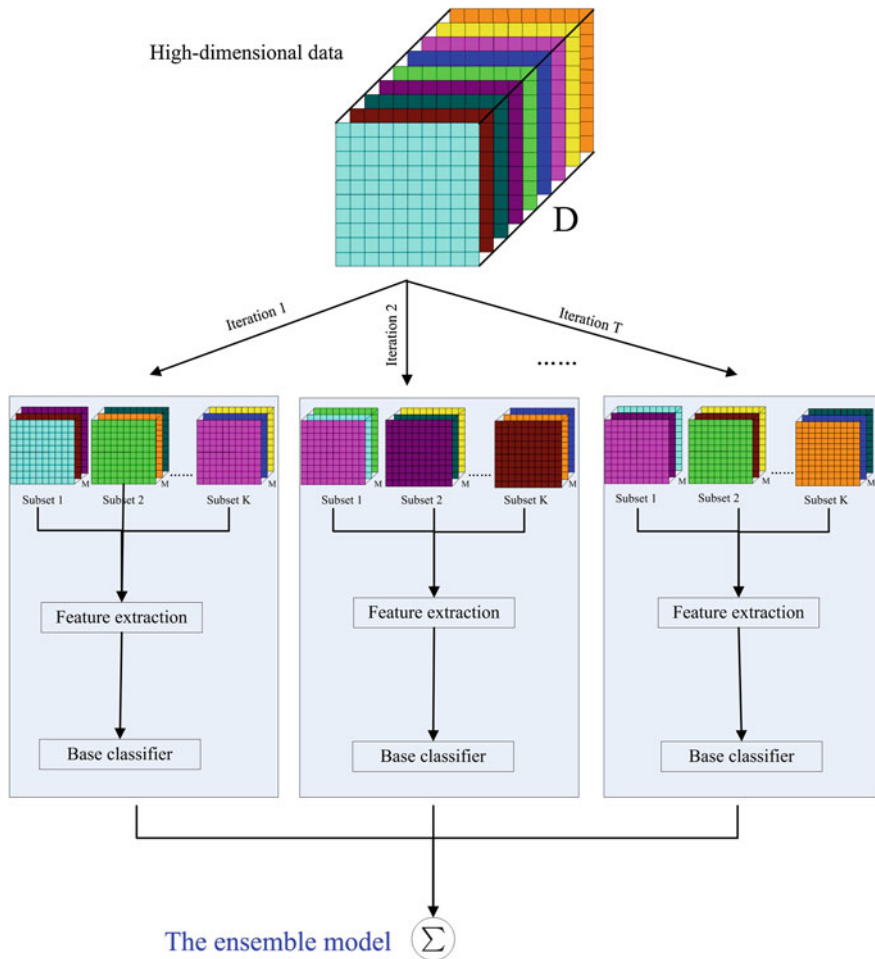


Fig. 6.3 Illustration of the rotation-based ensemble

- the input feature space is divided into K disjoint subspaces.
- feature extraction is performed on each subsets with the bootstrapped samples of 75 % size of $\{\mathbf{X}, Y\}$.
- the new training data, which is obtained by rotating the original training samples, is applied to the individual classifier.
- the individual classification results are combined using majority voting rule.

The strong performance is attributed to a simultaneous improvement of (1) diversity within the ensemble, obtained by the use of feature extraction on training data and (2) accuracy of the base classifiers, by keeping all extracted features in the training data [66, 72].

It is essential to notice step 5 in rotation-based ensemble presented in Algorithm 1, the sample size $\mathbf{X}'_{t,j}$ is selected smaller than $\mathbf{X}_{t,j}$ due to two reasons: one is to avoid obtaining the same coefficients when the same features are chosen and the other is to enhance the diversity within the ensemble [72].

Given the importance of the choice regarding the algorithm for feature extraction and the base classifier in rotation-based ensemble, several alternatives are considered in this study. The detailed feature extraction methods and base classifier can be found in the following section.

Algorithm 1 Rotation-based ensemble

Require: $\{\mathbf{X}, Y\} = \{\mathbf{x}_i, y_i\}_{i=1}^n$: training samples, T : number of classifier, K : number of subsets (M : number of features in each subset), Γ : base classifier

1. For $t = 1 : T$
2. randomly split the features F into K subsets $F_{t,j}$
3. For $j = 1 : K$
4. select the corresponding features of $F_{t,j}$ to compose a new training features $\mathbf{X}_{t,j}$
5. select a new training samples $\mathbf{X}'_{t,j}$ using bootstrap algorithm, whose size is 75 % of the original size
6. transform $\mathbf{X}'_{t,j}$ to get the coefficients $v_{t,j}^{(1)}, \dots, v_{t,j}^{(M_k)}$
7. Endfor
8. sparse matrix R_t is composed of the above coefficients

$$R_t = \begin{bmatrix} v_{t,1}^{(1)}, \dots, v_{t,1}^{(M_1)} & 0 & \dots & 0 \\ 0 & v_{t,2}^{(1)}, \dots, v_{t,2}^{(M_2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & v_{t,j}^{(1)}, \dots, v_{t,j}^{(M_K)} \end{bmatrix}$$

9. rearrange R_t to R_t^a with respect to the original feature set
10. obtain the new training samples $\{\mathbf{X}R_t^a, Y\}$
11. build classifier Γ_t using $\{\mathbf{X}R_t^a, Y\}$
12. Endfor

Ensure: the class label of given sample \mathbf{x} predicted by multiple classifier

$$\Gamma^*(\mathbf{x}) = \arg \max_{y \in \phi} \sum_{t=1}^T I(\Gamma_t(\mathbf{x}R_t^a) = y)$$

$I(a = b)$ equals to 1 when a equals to b , otherwise equals to 0

6.4 Two Implementations of Rotation-Based Ensemble

6.4.1 Rotation Forest

Decision trees are often used for the multiple classifier system, especially for the rotation-based ensembles, because it is sensitive and fast. In this chapter, we adapt Classification and Regression Tree (CART) to construct Rotation Forest (RoF).

CART is a nonparametric decision tree learning technique, which can be both used for classification and regression. Decision trees are formed by a collection of

rules based on variables in the modeling dataset: (1) rules based on variables's values are selected to get the best split to differentiate observations based on the dependent variable, (2) once a rule is selected and a node is split into two, the same process is applied to each 'child' node. (3) splitting stops when CART detects no further gain can be made, or some preset stopping rules are met. Each branch of the tree ends in a terminal node. Each observation falls into exactly one terminal node, and each terminal node is uniquely defined by a set of rules.

Both unsupervised and supervised feature extraction methods are applied to Rotation Forest. Principal Component Analysis (PCA) is the most popular linear unsupervised feature extraction method, which can keep the most information in a few components in terms of variance. Though Cheriadat and Bruce provide theoretical and experimental analysis to demonstrate that PCA is not optimal for dimensionality reduction in target detection and classification of hyperspectral data, PCA are still competitive for the purpose of classification because of its low complexity and the absence of parameters [16, 24].

Linear Discriminant Analysis(LDA) is the best-known supervised feature extraction approaches. But this method has the limitation: for C class classification problem, it can extract at maximum $C - 1$ features [18, 54]. That means in Rotation Forest, we should define the value of C is greater than K . In order to solve the problem, we adapt Local Fisher Discriminant Analysis (LFDA) instead of LDA. LFDA effectively combines the ideas of LDA and Locality Preserving Projection (LPP), which leads to both maximize between-class separability and preserve with-class local structure [80]. It can be viewed as the following eigenvalue decomposition problem:

$$\mathbf{S}^{lb} \mathbf{v} = \lambda \mathbf{S}^{lw} \mathbf{v} \quad (6.1)$$

where, \mathbf{v} is an eigenvector and λ is the eigenvalue corresponding to \mathbf{v} . \mathbf{S}^{lb} and \mathbf{S}^{lw} denote the local between-class and within-class scatter matrix. LFDA wants to find an eigenvector matrix that maximize the local between-class scatter in the embedding space while minimize the local within-class scatter in the embedding space. \mathbf{S}^{lb} and \mathbf{S}^{lw} can be defined:

$$\mathbf{S}^{lb} = \frac{1}{2} \sum_{i,j=1}^n \omega_{i,j}^{lb} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top \quad (6.2)$$

$$\mathbf{S}^{lw} = \frac{1}{2} \sum_{i,j=1}^n \omega_{i,j}^{lw} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top \quad (6.3)$$

where, ω^{lb} and ω^{lw} are the weight matrices with:

$$\omega_{i,j}^{lb} = \begin{cases} A_{i,j} \left(\frac{1}{n} - \frac{1}{n_{y_i}} \right) & y_i = y_j \\ \frac{1}{n} & y_i \neq y_j \end{cases} \quad (6.4)$$

$$\omega_{i,j}^{lw} = \begin{cases} \frac{A_{i,j}}{n_{y_i}} & y_i = y_j \\ 0 & y_i \neq y_j \end{cases} \quad (6.5)$$

where, $A_{i,j}$ is the affinity value between \mathbf{x}_i and \mathbf{x}_j in the local space.

$$A_{i,j} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\sigma_i \sigma_j}\right) \quad (6.6)$$

$$\sigma_i = \|\mathbf{x}_i - \mathbf{x}_i^e\| \quad (6.7)$$

where, \mathbf{x}_i^e is the e -th nearest neighbor of \mathbf{x}_i , n_{y_i} is the number of labeled samples in class $y_i \in \{1, 2, 3, \dots, C\}$.

6.4.2 Rotation SVM

SVM classifier has shown better classification performance for high-dimensional data than other classifier. SVM is very stable that small changes in the training set cannot produce very different SVM classifiers.

Therefore, it is difficult to get an ensemble of multiple SVM that perform better than a single SVM using the state of the art ensemble methods. Thus, we hope to introduce more diversity into SVM. In [52], diversity is analyzed for Random Projections (RP) with and without splitting into group of attributes. Therefore, we introduce Random Projection (RP) into rotation-based SVM in order to promote the diversity within the ensemble.

RP obtains the rotation matrix using simply random number. Unlike other feature extraction methods such as PCA, RP can get a projected space which is bigger than the original. Two types of RP are used in this chapter [1]:

1. Gaussian. Each value in transformation matrix comes from a Gaussian distribution (mean 0 and standard deviation).
2. Sparse. The entry values are $\sqrt{3} \times \alpha$, where, α is a random number taking the following value: -1 with the probability $1/6$, 0 with the probability $2/3$ and $+1$ with probability $1/6$.

6.5 Experimental Results and Analysis

6.5.1 Experimental Setup

In this section, we present the results that we obtained with rotation-based ensemble on different types of images. Two airborne hyperspectral images are used to evaluate

Rotation Forest (RoF). An airborne hyperspectral and a multi-date remote sensing images are applied to test the performance of Rotation SVM (RoSVM). The descriptions of the data are detailed in the following two subsections. Overall accuracy (OA), average accuracy (AA), and class-specific accuracy are used to evaluate the efficiency of RoF and RoSVM.

Popular ensemble methods, including Bagging [9], AdaBoost [28] and Random Forest (RF) [10] are added to be compared with Rotation Forest. The performance achieved by Rotation Forest is illustrated using the following design:

- Number of features in each subset: $M = 10$;
- Number of classifiers in the ensemble: $L = 10$;
- Feature extraction method: PCA [42] and LFDA [80];

we employed RoF-PCA and RoF-LFDA as the abbreviations of Rotation Forest with PCA and LFDA.

Gaussian RBF kernel is adopt in the SVM. In order to reduce the computational time in the ensembles of SVM, we used the fixed parameters in SVM. Random Projection-based ensemble is added to compare with RoSVM using RP projections. Two sizes of projected space dimension have been tested (100 and 150 %). The configurations of 150 % size are denoted as RoSVM or RP 150 %. The performance achieved by RoSVM is illustrated using the following designs:

- Number of features in each subset: $M = 10$;
- Number of classifiers in the ensemble: $L = 10$;
- Feature extraction method: Random Projection (RP) with *Gaussian* and *Sparse*;
- Base classifier: SVM.

In the following experiments, we employed RP and RoSVM as the abbreviations of Random Projection-based ensemble and rotation-based SVM ensemble.

6.5.2 *Rotation Forest*

6.5.2.1 **Indiana Pines AVIRIS Image**

The first hyperspectral image is recorded by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indiana Pines in Northwestern Indiana, USA. The image is composed of 145×145 pixels, and the spatial resolution is 20 m per pixel. This image is a classical benchmark to validate the accuracy of hyperspectral image analysis algorithms and constitutes a challenging problem due to the significant presence of mixed pixels in all available classes and also because of the unbalanced number of available labeled pixels per class. The three-band color composite and ground truth of AVIRIS image can be seen in Fig. 6.4. We have chosen 20 pixels of each class from the available ground truth (a total size of 320 pixels) as the training set.

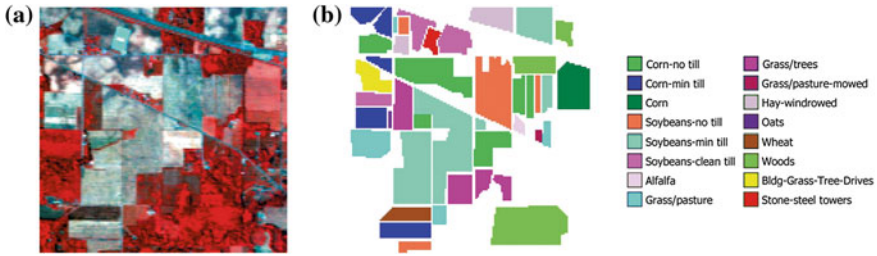


Fig. 6.4 **a** Three-band color composite of AVIRIS image. **b** Ground truth: Corn-no till, corn-min till, corn, soybean-no till, soybeans-min till, soybeans-clean till, alfalfa, grass/pasture, grass/trees, grass/pasture-mowed, hay-windrowed, oats, wheat, woods, bldg-grass-tree-drives, stone-steel towers

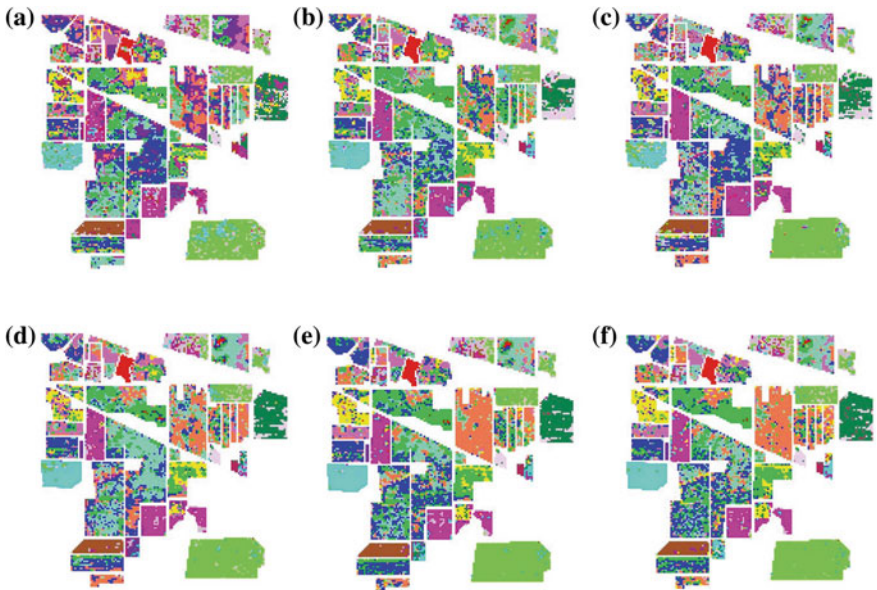


Fig. 6.5 Classification results of Indiana Pines AVIRIS image. Different color represents the different class. The color of the classes can be found in Fig. 6.4. **a** CART. **b** Bagging. **c** AdaBoost. **d** RF. **e** RoF-PCA. **f** RoF-LFDA

Table 6.4 shows the classification accuracies (OA %) obtained by the Rotation Forest approaches as well as other algorithms using different training samples. We highlight the highest accuracies of each case in bold font. From Table 6.4, it can be seen that RoF-PCA and RoF-LFDA achieve better results than other ensemble approaches (Bagging, Adaboost, and RF). Compared to Bagging, AdaBoost and RF, Rotation Forest can promote the diversity and improve the accuracy of individual classifier within the ensemble. Therefore, in most cases, Rotation Forest is superior to Bagging, AdaBoost, and Random Forest. Our experimental results are compati-

Table 6.4 Overall, average and class-specific accuracies of the Indiana Pines AVIRIS image

Class	Train	Test	CART	Bagging	AdaBoost	RF	RoF-PCA	RoF-LFDA
Alfalfa	20	54	77.78	79.63	79.63	87.04	88.89	88.89
Corn-no till	20	1434	32.15	52.09	37.66	37.34	50.28	50.84
Corn-min till	20	834	37.89	37.53	45.92	48.08	61.51	54.8
Bldg-grass-tree-drives	20	234	44.44	50.85	44.02	47.44	70.51	78.21
Grass/pasture	20	497	46.08	71.23	46.48	63.38	76.46	78.47
Grass/trees	20	747	57.43	79.92	84.47	73.9	73.63	76.44
Grass/pasture-mowed	20	26	88.46	88.46	92.31	92.31	96.15	92.31
Corn	20	489	62.99	49.69	58.69	81.6	86.09	84.05
Oats	20	20	30	90	75	100	100	100
Soybeans-no till	20	968	30.48	42.25	49.38	53.71	76.55	76.65
Soybeans-min till	20	2468	23.3	35.13	31	45.79	31.69	30.88
Soybeans-clean till	20	614	28.66	31.92	34.2	43.81	47.56	51.79
Wheat	20	212	86.79	88.68	86.79	92.45	94.34	91.98
Woods	20	1294	69.09	76.82	83	83.77	91.19	89.49
Hay-windrowed	20	380	48.16	45.79	46.84	55.53	48.42	48.42
Stone-steel towers	20	95	74.74	96.84	97.89	95.79	93.68	94.74
OA			41.44	51.87	50.54	56.97	60.88	60.60
AA			52.4	63.55	63.55	68.87	74.18	74.25

ble with the theoretical analysis. For instance, CART, Bagging, Adaboost and RF acquired an OA of 41.44, 51.87, 50.54 and 56.97 %, respectively. RoF-PCA and RoF-LFDA respectively increased the OA to 60.88 and 60.6 %, while the AA of RoF-PCA and RoF-LFDA were improved to 23.78 and 23.85 % percentage points compared to CART. The OA of RoF-PCA is slightly higher than the one of RoF-LFDA. But there is no significantly difference between the two classification results according to McNemar test. Nine of sixteen class-specific accuracies is improved by RoF-PCA and RoF-LFDA.

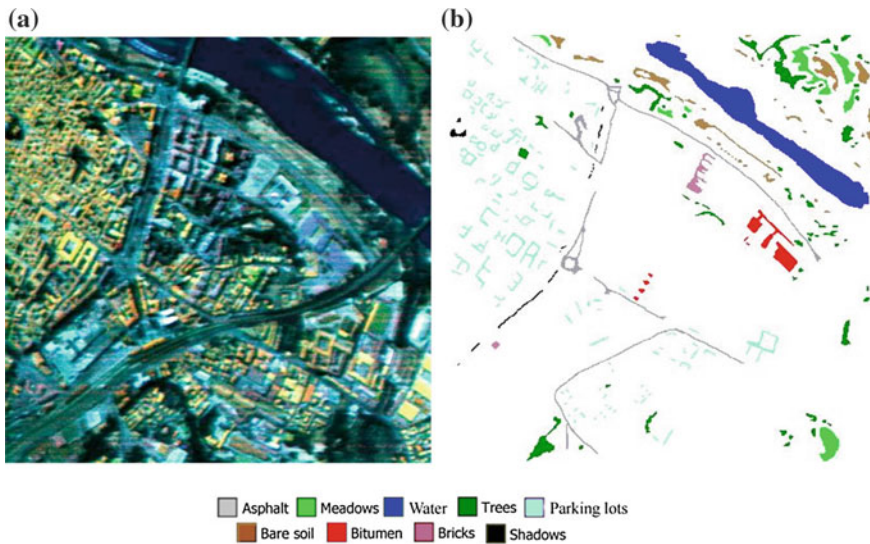
The classification results of Indiana Pines AVIRIS image are shown in Fig. 6.5. The classification map for the CART classifier was very noisy because CART is not a promising classifier for high-dimensional data. Compared to the reference data presented in Fig. 6.4b, all the ensemble methods produced more corrected classification results than CART. If we carefully look at the reference image, particularly, the area of *Soybean-no till*, this region is almost correctly classified by RoF-PCA and RoF-LFDA, whereas it is classified as *Corn-min till* and *Corn-no till* by other classifiers.

6.5.2.2 Pavia Center DAIS Image

The second image was acquired by the DAIS sensor at 1500 m flight altitude over the city of Pavia, Italy. The image (seen in Fig. 6.6) has a size of 400×400 pixels, with ground resolution of 5 m. The 80 data channels recorded by this spectrometer were used for this experiment. Nine land cover classes of interest are considered, which are detailed in Table 6.5, with the number of labeled samples for each class.

Table 6.5 Overall, average and class-specific accuracies of the Pavia Center DAIS image

Class	Train	Test	CART	Bagging	AdaBoost	RF	RoF-PCA	RoF-LFDA
Water	10	4281	98.11	96.15	90.66	100	100	100
Trees	10	2424	56.89	67.12	63.74	88.7	87.83	91.3
Meadows	10	1251	97.44	97.76	97.2	99.52	99.12	99.12
Bricks	10	2237	74.17	77.08	72.68	65.02	84.27	80.2
Soil	10	1475	50.29	60.39	65.38	76.7	74.77	74.71
Asphalt	10	1704	77.7	77.7	83.28	77.35	91.99	97.91
Bitumen	10	685	68.22	63.66	76.62	83.33	94.95	93.74
Parking lot	10	287	70.22	87.01	91.09	78.1	86.28	88.32
Shadows	10	241	86.72	66.39	85.06	92.95	95.85	89.63
OA			76.71	79.09	79.6	87.67	91.72	91.8
AA			75.53	77.03	80.63	84.63	90.56	90.55

**Fig. 6.6** a Three-band color composite of DAIS image. b Ground truth

The global accuracies and class-specific accuracies of the Pavia Center DAIS image are reported in Table 6.5. The classification results achieved by the ensemble classifiers are similar with the ones of AVIRIS image. Regarding the overall accuracies, Rotation Forest with different feature extraction algorithms are all superior to other approaches under comparison. RoF-LFDA yields the highest OA (91.8%). The accuracies of class *Bricks*, *Asphalt* and *Bitumen* are significantly improved by the Rotation Forest ensemble classifiers. The classification results of Pavia Center DAIS image are shown in Fig. 6.7.

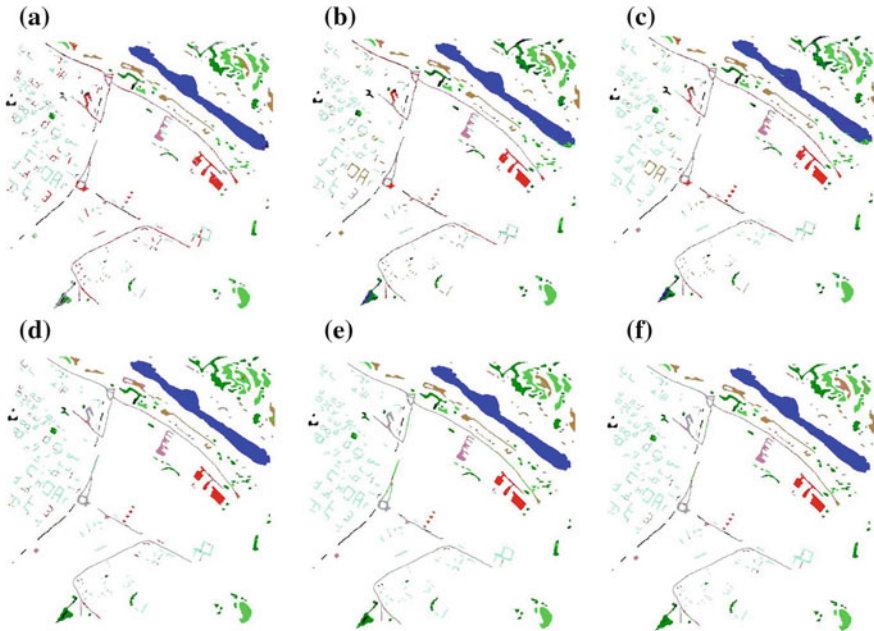


Fig. 6.7 Classification results of Pavia Center DAIS image. **a** CART. **b** Bagging. **c** AdaBoost. **d** RF. **e** RoF-PCA. **f** RoF-LFDA

6.5.2.3 Sensitivity of the Parameters

Ensemble size (T) and the number of features in a subset (M) are the key parameters of Rotation Forest, which are indicators 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 T is fixed to 10. Fig. 6.8 shows the OA (%) using different number of T and M obtained from AVIRIS and DAIS images. For AVIRIS image, the classification results are improved when the values of T and M increased. RoF-PCA is superior to RoF-LFDA. For DAIS image, the OAs are improved with the increasement of T . The general trend of different values of M is not obvious.

6.5.2.4 Discussion

Based on the above classification results, we identify that the incorporation of multiple classifiers has shown great improvement for the classification of high-dimensional data. In order to make MCS effective, we should enforce the diversity by the manipulation of training sets. Bagging and Boosting aim at changing the distribution of the training samples to obtain the different training set. Random subspace method

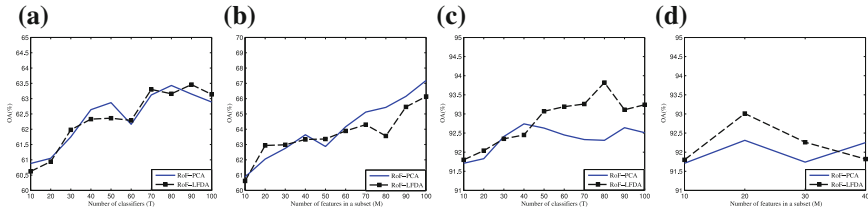


Fig. 6.8 Effects of varying parameters on the performance of rotation forest. Indiana Pines AVIRIS image. **a** Sensitivity for change of T ($M = 10$). **b** Sensitivity for change of M ($T = 10$). Pavia Center DAIS image. **c** Sensitivity for change of T ($M = 10$). **d** Sensitivity for change of M ($T = 10$)

constructs several classifier by random selecting the subset of the features. It is very useful for the classification problem where the number of features is much larger than the number of training samples. Random subspace method is a generalization of the Random Forest algorithm, whereas Random Forest is composed of decision trees. Rotation Forest tries to create the individual classifiers that are both diverse and accurate, each based on a different axis rotation of attributes. To create different training set, the features are randomly split into a given number of subsets and feature extraction is applied to each subset. Decision trees is very sensitive to the rotation of axis. In this chapter, we select CART to construct Rotation Forest. Rotation Forest can promote more diversity than Bagging, AdaBoost and Random Forest. Therefore, it can produce more accurate results than Bagging, AdaBoost and Random Forest. An important issue of Rotation Forest is the selection of the parameters (T and M). A larger value of T will often increase the accuracy and also increase the computation time. The optimal value of M is hard to determine. Different datasets achieve the highest accuracy with different value of M . The computation time of Rotation Forest approaches is longer than those of Bagging, AdaBoost and Random Forest. But the computation complexity of Rotation Forest is much less than the one of the strong classifier of high-dimensional data, such as SVM.

6.5.3 Rotation SVM

6.5.3.1 Indiana Pines AVIRIS Image

Table 6.6 shows overall, average and class-specific accuracies using different version of rotation-based SVMs. We highlight the highest accuracies of each case in bold font. It can be seen that RoSVM achieve the better results than RP, because RoSVM can provide more diversity than RP. For this dataset, RP with Gaussian is superior to the one of Sparse. By employing a slightly higher size of a projected space, the results of RP is improved but RoSVM yields bad results. The corresponding results are shown in Fig. 6.9. We have studied the impacts of T and M in RoSVM. The sensitivity of performance using different T and M is not obvious.

Table 6.6 Overall, average and class-specific accuracies of the Indiana Pines AVIRIS image

Class	SVM		RP		RoSVM			
	Gaussian	Sparse	Gaussian	Sparse	Gaussian	Gaussian 150 %	Sparse	Sparse 150 %
Alfalfa	85.19	90.74	88.89	94.44	100	94.44	94.44	88.89
Corn-no till	23.78	62.41	66.46	57.32	57.74	56	51.88	62.76
Corn-min till	48.44	52.39	57.19	51.68	57.55	62.23	56.36	55.88
Bldg-Grass-Tree-Drives	78.63	80.77	79.06	73.50	80.77	70.51	84.19	64.53
Grass/pasture	79.07	78.07	81.09	84.51	88.33	87.32	88.13	89.74
Grass/trees	77.24	77.78	85.14	88.09	91.97	89.56	92.24	93.17
Grass/pasture-mowed	92.31	100	100	96.15	100	96.15	100	96.15
Corn	85.48	68.30	90.39	82	95.09	92.64	92.23	96.11
Oats	75	100	100	95	100	100	100	90
Soybeans-no till	45.35	57.85	58.99	50.20	65.08	60.54	60.12	60.02
Soybeans-min till	52.92	45.83	36.47	40.44	49.31	52.15	48.87	46.31
Soybeans-clean till	39.74	57.98	64.01	62.38	59.45	47.23	58.14	41.86
Wheat	98.58	98.11	99.06	95.75	98.58	97.17	99.53	99.06
Woods	92.74	84.18	93.12	79.6	90.57	83.93	92.35	87.86
Hay-windrowed	28.16	57.89	51.58	64.73	57.63	63.94	65.53	59.21
Stone-steel towers	96.84	92.63	87.37	93.68	96.84	100	96.84	88.42
OA	57.87	63.39	65.1	62.07	68.42	66.87	66.85	66.15
AA	68.72	75.31	77.43	75.59	80.56	78.36	80.05	76.25



Fig. 6.9 Classification results of Indiana Pines AVIRIS image. **a** SVM. **b** RP Gaussian. **c** RP Gaussian 150 %. **d** RP Sparse. **e** RP Sparse 150 %. **f** RoSVM Gaussian. **g** RoSVM Gaussian 150 %. **h** RoSVM Sparse. **i** RoSVM Sparse 150 %

6.5.3.2 CHRIS Multi-date Images

The second high-dimensional data is the three dates of Compact High-Resolution Imaging Spectrometer (CHRIS) images acquired by the Project for On-Board Autonomy (PROBA)-1 satellite with spatial resolution of 21 m/pixel. The total number of spectral bands is 54. More details about CHRIS image can be seen in [23]. Training samples contains 2,297 samples and test data includes 1,975 samples with 11 classes.

The flowchart of RoSVM for classifying CHRIS image is the same as the previous AVIRIS dataset. Single SVM achieved the accuracy of 84.05 %. All the methods based on RP and RoSVM ensemble can generate the better accuracies than a single SVM. In particular, RoSVM ensembles are slightly superior to RP ensembles because they enforce the diversity by applying RP to the subsets of the features. RoSVM with Sparse RP gains the highest the overall accuracy.

6.5.3.3 Discussion

SVM is a stable classifier, so it is hard to generate different individual SVM classifiers using the common manipulation ways. Therefore, we should introduce more diversity to construct the diverse individual SVM classifiers. In this chapter, we adapt Random Projection methods to produce diverse SVM classifiers. Two sizes of projected space dimension have been tested. Experimental results indicated that RoSVM ensemble outperform RP ensembles. The main drawback of RoSVM is the computational complexity, especially for large training samples. The sensitivity of performance using different M is not obvious.

6.6 Conclusion

In this chapter, we first presented a review of MCS approaches with special focus on applications of high-dimensional data. Recently rotation-based ensemble classifiers were applied to high-dimensional data. They consist in splitting the feature set into several subsets, running feature extraction algorithms separately on each subset and then reassembling a new extracted feature set while keeping all the components. CART Decision Tree and SVM classifiers are used as the base classifier. Different splits of the feature set lead to different rotations. Thus diverse classifiers are obtained. We take into account both diversity and accuracy. Rotation Forest using PCA, LFDA, Rotation SVM using RP are used to classify high-dimensional data.

Experimental results have shown that rotation-based ensemble methods (both DT and SVM) outperform classical ensemble methods such as Bagging, AdaBoost, Random Forest in terms of accuracies. The key parameters are also explored in this chapter. Future studies will be devoted to the integration of rotation-based ensemble classifiers with other ensemble approaches, the selection of an optimized Decision Tree model, and the use of other effective feature extraction algorithms.

References

1. Achlioptas, D (2001) Database-friendly random projections. In: Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems, New York, NY, USA, 2001, pp 274–281
2. Aksela M, Laaksonen J (2006) Using diversity of errors for selecting members of a committee classifier. *Pattern Recogn* 39(4):608–623

3. Bakos KL, Gamba P (2011) Hierarchical hybrid decision tree fusion of multiple hyperspectral data processing chains. *IEEE Trans Geosci Remote Sens* 49(1–2):388–394
4. Benediktsson JA, Chanussot J, Fauvel M (2007) Multiple classifier systems in remote sensing: from basics to recent developments. *Lect Notes Comput Sci* 4472:501–512
5. Benediktsson JA, Sveinsson JR (1997) Hybrid consensus theoretic classification. *IEEE Trans Geosci Remote Sens* 35(4):833–843
6. Benediktsson JA, Swain PH (1992) Consensus theoretic classification methods. *IEEE Trans Syst Man Cybern* 22(4):688–704
7. Biggio B, Fumera G, Roli F (2010) Multiple classifier systems for robust classifier design in adversarial environments. *J Mach Learn Cybern* 1:27–41
8. Braun AC, Weidner U, Hinz S (2012) Classification in high-dimensional feature spaces assessment using SVM, IVM and RVM with focus on simulated EnMap data. *IEEE J Sel Top Appl Earth Observations Remote Sens* 5(2):436–443
9. Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140
10. Breiman L (2001) Random forest. *Mach Learn* 45(1):5–32
11. Briem G, Benediktsson J, Sveinsson J (2002) Multiple classifiers applied to multisource remote sensing data. *IEEE Trans Geosci Remote Sens* 40(10):2291–2299
12. Brown G, Wyatt J, Harris R, Yao X (2005) Diversity creation methods: a survey and categorisation. *Inf Fusion* 6(1):5–20
13. Bryll RK, Gutierrez-Osuna R, Quek FKH (2003) Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets. *Pattern Recogn* 36(6):1291–1302
14. Chan JC, Paelinckx D (2008) Evaluation of Random Forest and AdBoost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sens Environ* 112(6):2999–3011
15. Chandra A, Yao X (2006) Evolving hybrid ensembles of learning machines for better generalisation. *Neurocomputing* 69(7–9):686–700
16. Cheriyyadat A, Bruce LM (2003) Why principal component analysis is not an appropriate feature extraction method for hyperspectral data. In: *Proceedings of IEEE geoscience and remote sensing symposium (IGARSS)*, Toulouse, France, 2003, pp 3420–3422
17. Cunningham P, Carney J (2000) Diversity versus quality in classification ensembles based on feature selection. In: *11th European conference on machine learning*, Barcelona, Spain, 2000, pp 109–116
18. Dell’Acqua F, Gamba P, Ferari A, Palmason BJA, Arnason K (2004) Exploiting spectral and spatial information in hyperspectral urban data with high resolution. *IEEE Geosci Remote Sens Lett* 1(4):322–326
19. Dietterich TG, Bakiri G (1995) Solving multiclass learning problems via error-correcting output codes. *J Artif Intell Res* 2(1):263–286
20. Doan HTX, Foody GM (2007) Increasing soft classification accuracy through the use of an ensemble of classifiers. *Int J Remote Sens* 28(20):4609–4623
21. Du P, Xia J, Zhang W, Tan K, Liu Y, Liu S (2012) Multiple classifier system for remote sensing image classification: a review. *Sensors* 12(4):4764–4792
22. Du P, Zhang W, Xia J (2011) Hyperspectral remote sensing image classification based on decision level fusion. *Chin Optics Lett* 9(3):031002–031004
23. Duca R, Frate FD (2008) Hyperspectral and multiangle CHRIS-PROBA images for the generation of land cover maps. *IEEE Trans Geosci Remote Sens* 10:2857–2866
24. Fauvel M, Chanussot J, Benediktsson JA (2009) Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas. *EURASIP J Adv Signal Process* 2:1–15
25. Fleiss J (1981) *Statistical methods for rates and proportions*. Wiley, New York
26. Foody GM (2009) Classification accuracy comparison: hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. *Remote Sens Environ* 113(8):1658–1663
27. Foody GM, Boyd DS, Sanchez-Hernandez C (2007) Mapping a specific class with an ensemble of classifiers. *Int J Remote Sens* 28:1733–1746

28. Freund Y, Schapire RE (1996) Experiments with a new Boosting algorithm. In: International conference on machine learning, Bari, Italy, 1996, pp 148–156
29. Gao J, Liang F, Fan W, Sun Y, Han J (2013) A graph-based consensus maximization approach for combining multiple supervised and unsupervised models. *IEEE Trans Knowl Data Eng* 25(1):15–28
30. Giacinto G, Roli F (2001) Design of effective neural network ensembles for image classification. *Image Vis Comput J* 19(9/10):697–705
31. Giacinto G, Roli F, Fumera G (2000) Selection of image classifiers. *Electron Lett* 36:420–422
32. Giacinto G, Roli F, Vernazza G (1997) Comparison and combination of statistical and neural network algorithms for remote-sensing image classification. *Neurocomputation in remote sensing data analysis*. Springer, Berlin, pp 117–124
33. Gislason PO, Benediktsson JA, Sveinsson JR (2006) Random forests for land cover classification. *Pattern Recogn Lett* 27(4):294–300
34. Gmez-Chova L, Fernandez-Prieto D, Calpe-Maravilla J, Soria-Olivas E, Vila-Francis J, Camps-Valls G (2006) Urban monitoring using multi-temporal SAR and multi-spectral data. *Pattern Recogn Lett* 27(4):234–243
35. Guyon I, Gunn S, Nikravesh M, Zadeh L (2006) Feature extraction: foundations and applications. Springer, New York
36. Ham J, Chen Y, Crawford MM, Ghosh J (2005) Investigation of the random forest framework for classification of hyperspectral data. *IEEE Trans Geosci Remote Sens* 43(3):492–501
37. Hansen LK, Salamon P (1990) Neural network ensembles. *IEEE Trans Pattern Anal Mach Intell* 12(10):993–1001
38. He Z, Xu X, Deng S (2005) A cluster ensemble method for clustering categorical data. *Inf Fusion* 6(2):143–151
39. Ho TK (1998) The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 20(8):832–844
40. Hughes G (1968) On the mean accuracy of statistical pattern recognizers. *IEEE Trans Inf Theory* 14(1):55–63
41. Iam-On N, Boongoen T, Garrett S, Price C (2011) A link-based approach to the cluster ensemble problem. *IEEE Trans Pattern Anal Mach Intell* 33(12):2396–2409
42. Richards JA, Jia X (2006) Remote sensing digital image analysis: an introduction. Springer, New York
43. Kang H-J, Lee S-W (2000) An information-theoretic strategy for constructing multiple classifier systems. In: International conference on pattern recognition (ICPR), Barcelona, Spain, 2000, pp 2483–2486
44. Kawaguchi K, Nishii R (2007) Hyperspectral image classification by bootstrap AdaBoost with random decision stumps. *IEEE Trans Geosci Remote Sens* 45(11–12):3845–3851
45. Kohavi R, Wolpert DH (1996) Bias plus variance decomposition for zero-one loss functions. In: 13th international conference on machine learning (ICML), Bari, Italy, 1996, pp 275–283
46. Kong Z, Cai Z (2007) Advances of research in fuzzy integral for classifiers' fusion. In: Proceedings of the eighth ACIS international conference on software engineering, artificial intelligence, networking, and parallel/distributed computing, Washington, DC, USA pp 809–814
47. Krogh A, Vedelsby J (1995) Neural network ensembles, cross validation, and active learning. *Adv Neural Inf Process Syst* 7:231–238
48. Kumar S, Ghosh J, Crawford MM (2002) Hierarchical fusion of multiple classifiers for hyperspectral data analysis. *Pattern Anal Appl* 5:210–220
49. Kuncheva L, Whitaker C, Shipp C, Duin R (2000) Is independence good for combining classifiers? In: International conference on pattern recognition (ICPR), 2000, p 2168
50. Kuncheva LI (2004) Combining pattern classifiers: methods and algorithms. Wiley-Interscience, New Jersey
51. Kuncheva LI (2005) Diversity in multiple classifier systems. *Inf Fusion* 6(1):3–4
52. Kuncheva LI, Rodríguez JJ (2007) An experimental study on rotation forest ensembles. In: Proceedings of the 7th international workshop on multiple classifier systems, Prague, Czech Republic, 2007, pp 459–468

53. Kuncheva LI, Whitaker CJ (2003) Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Mach Learn* 51(2):181–207
54. Landgrebe DA (1984) *Signal theory methods in multispectral remote sensing*. Wiley, New York
55. Lawrence R, Bunna A, Powellb S, Zambon M (2004) Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sens Environ* 90(3):331–336
56. Lu D, Weng Q (2007) A survey of image classification methods and techniques for improving classification performance. *Int J Remote Sens* 28(5):823–870
57. Margineantu DD, Dietterich TG (1997) Pruning adaptive boosting. In: *Proceedings of the fourteenth international conference on machine learning*, San Francisco, CA, USA, 1997, pp 211–218
58. Martinez-Munoz G, Suarez A (2005) Switching class labels to generate classification ensembles. *Pattern Recogn* 38(10):1483–1494
59. Melgani F, Bruzzone L (2004) Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans Geosci Remote Sens* 42(8):1778–1790
60. Michail P, Benediktsson JA, Ioannis K (2002) The effect of classifier agreement on the accuracy of the combined classifier in decision level fusion. *IEEE Trans Geosci Remote Sens* 39(11):2539–2546
61. Mojaradi B, Abrishami-Moghaddam H, Zoj M, Duin R (2009) Dimensionality reduction of hyperspectral data via spectral feature extraction. *IEEE Trans Geosci Remote Sens* 47(7):2091–2105
62. Moon H, Ahn H, Kodell RL, Baek S, Lin C-J, Chen JJ (2007) Ensemble methods for classification of patients for personalized medicine with high-dimensional data. *Artif Intell Med* 41(3):197–207
63. Moreno-Seco F, Inesta J, Ponce De Leon P, Mico L (2006) Comparison of classifier fusion methods for classification in pattern recognition tasks. *Lect Notes Comput Sci* 4109:705–713
64. Navalgund RR, Jayaraman V, Roy PS (2007) Remote sensing applications: an overview. *Current* 93(12):1747–1766
65. Nemmour H, Chibani Y (2006) Multiple support vector machines for land cover change detection: an application for mapping urban extensions. *ISPRS J Photogrammetry Remote Sens* 61(2):125–133
66. Ozcift A, Gulden A (2011) Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. *Comput Methods Programs Biomed* 104(3):443–451
67. Pal M, Foody GM (2010) Feature selection for classification of hyperspectral data by SVM. *IEEE Trans Geosci Remote Sens* 48(5):2297–2307
68. Partridge D, Krzanowski W (1997) Software diversity: practical statistics for its measurement and exploitation. *Inf Softw Technol* 39(10):707–717
69. Plaza A, Martinez P, Plaza J, Perez R (2005) Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations. *IEEE Trans Geosci Remote Sens* 43(3):466–479
70. Rahman A, Fairhurst M (1999) Serial combination of multiple experts: a unified evaluation. *Pattern Anal Appl* 2:292–311
71. Ranawana R, Palade V (2006) Multi-classifier systems: review and a roadmap for developers. *Int J Hybrid Intell Syst* 3(1):1–41
72. Rodriguez JJ, Kuncheva LI (2009) Rotation forest: a new classifier ensemble method. *IEEE Trans Pattern Anal Mach Intell* 28(10):1619–1630
73. Rodriguez-Galianoa VF, Ghimireb B, Roganb J, Chica-Olmoa M, Rigol-Sanchez JP (2011) An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J Photogrammetry Remote Sens* 67:93–104
74. Rokach L (2010) *Pattern classification using ensemble methods*. World Scientific, Singapore
75. Ruitenbeek F, Debba P, Meer F, Cudahy T, Meijde M, Hale M (2006) Mapping white micas and their absorption wavelengths using hyperspectral band ratios. *Remote Sens Environ* 102(3–4):211–222

76. Ruta D, Gabrys B (2000) An overview of classifier fusion methods. *Comput Inf Syst* 7(1):1–10
77. Schölkopf B, Platt JC, Shawe-Taylor J, Smola AJ, Williamson RC (2001) Estimating the support of a high-dimensional distribution. *Neural Comput* 13(7):1443–1471
78. Smits PC (2002) Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection. *IEEE Trans Geosci Remote Sens* 40:801–813
79. Steele BM (2000) Combining multiple classifiers: an application using spatial and remotely sensed information for land cover type mapping. *Remote Sens Environ* 74(3):545–556
80. Sugiyama M (2007) Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *J Mach Learn Res* 27(8):1021–1064
81. Sun Q, Ye XQ, Gu WK (2000) A new combination rules of evidence theory (in chinese). *Acta Electronica Sinica* 8(1):117–119
82. Tsoumakas G, Partalas I, Vlahavas I (2009) An ensemble pruning primer. In: Okun O, Valentini G (eds) *Applications of supervised and unsupervised ensemble methods*. Springer, Berlin, pp 1–13
83. Vapnik VN (1995) *The nature of statistical learning theory*. Springer, Berlin
84. Waske B, Benediktsson JA, Arnason K, Sveinsson JR (2009) Mapping of hyperspectral aviris data using machine-learning algorithms. *Can J Remote Sens* 35:106–116
85. Waske B, Van Der Linden S, Benediktsson JA, Rabe A, Hostert P (2010) Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Trans Geosci Remote Sens* 48(7):2880–2889
86. Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. *IEEE Trans Evol Comput* 1(1):67–82
87. Woods K, Kegelmeyer WP, Bowyer K (1997) Combination of multiple classifiers using local accuracy estimates. *IEEE Trans Pattern Anal Mach Intell* 19(4):405–410
88. Wozniak M, Grana M, Corchado E (2014) A survey of multiple classifier systems as hybrid systems. *Inf Fusion* 16(1):3–17
89. Xia J, Du P, He X, Chanussot J (2014) Hyperspectral remote sensing image classification based on rotation forest. *IEEE Geosci Remote Sens Lett* 11:239–243
90. Xu L, Krzyzak A, Suen CY (1992) Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Trans Syst Man Cybern* 22(3):418–435
91. Yang J-M, Kuo B-C, Yu P-T, Chuang C-H (2010) A dynamic subspace method for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 48(7):2840–2853
92. Zeng H, Triosell HJ (2004) Dimensionality reduction in hyperspectral image classification. In: *IEEE International Conference on Image Processing (ICIP)*, Singapore, 2004, pp 913–916
93. Zhang Y (2010) Ten years of technology advancement in remote sensing and the research in the crc-agip lab in gce. *Geomatica* 64(2):173–189