CLASSIFICATION OF REMOTE SENSING DATA USING MARGIN-BASED ENSEMBLE METHODS

Samia Boukir1, Li Guo1,2, Nesrine Chehata1,3*

1G&E Laboratory (EA 4592), IPB / University of Bordeaux, 1 allée F. Daguin 33670 Pessac, France.
2CNRS - IMS Laboratory (UMR 5218), Bt A4, 351 Cours de la libération, 33402 Talence cedex, France.
3IRD, UMR LISAH (INRA-IRD-SupAgro), T- 1004 El Menzah, Tunisie.
E-mail: name.surname@ipb.fr

ABSTRACT
This work exploits the margin theory to design better ensemble classifiers for remote sensing data. The margin paradigm is at the core of a new bagging algorithm. This method increases the classification accuracy, particularly in case of difficult classes, and significantly reduces the training set size. The same margin framework is used to derive a novel ensemble pruning algorithm. This method not only highly reduces the complexity of ensemble methods but also performs better than complete bagging in handling minority classes. Our techniques have been successfully used for the classification of remote sensing data.

Index Terms—Bagging, ensemble margin, ensemble pruning, multiple classifier, remote sensing.

1. INTRODUCTION
Classification methods have been increasingly popular in the remote-sensing community for decades. Classification has indeed various environmental applications in earth-observing images [1, 2, 3]. It helps providing a land cover mapping at large scales. Hence, great efforts have been made in developing advanced classification approaches for improving the classification accuracy [4].

Some issues are specific to remote sensing data [5]. First, satellite images present a large amount of data, covering large areas. The ground truth is generally assessed on the field, thus the training data should be reduced. Secondly, the data can be highly imbalanced, especially in urban mapping. These challenging issues are tackled through a multiple classifier framework based on the margin paradigm.

Ensemble methods build a classification model by integrating multiple component learners [6]. One of the most successful ensemble methods is bagging [7]. It is made of bootstrap-inspired classifiers [8], produced by sampling with replacement from training instances, and uses them to get an aggregated classifier.

Ensemble margin is a key concept in ensemble learning [9]. Apart from our previous work [10, 5], which used the ensemble margin for feature selection, the margin has never been investigated in remote sensing. In this work, we exploit this concept to efficiently and effectively map remotely sensed data [5], at two learning levels: data level and classifier level. Our previous contribution [10, 5] having dealt with the feature level. At data level, an innovative bagging algorithm, exploiting low margin instances, is proposed to better handle huge and/or imbalanced datasets, commonly encountered in remote sensing. At classifier level, a novel ensemble pruning algorithm, also relying on lower margin instances, aims at reducing the number of components of the ensemble while maintaining its classification performance. This important issue has never been investigated yet in remote sensing despite the computational complexity in classifying remote sensing data. To our knowledge, ensemble methods (mainly Random Forests) have been used in remote sensing mainly as a classification tool for site characterization [11, 12, 13].

2. ENSEMBLE MARGIN
Ensemble margin [14] is a fundamental concept in ensemble learning. We use an alternative [15, 5] to the classical definition [14] of ensemble margin with an appealing property: it does not require the true class labels of instances. Thus, it is potentially more robust to noise as it is not affected by errors occurring on the class label itself. This unsupervised margin is particularly well-suited for remote sensing data which usually present a significant amount of mislabeled instances. It can be computed by equation 1, where $c_1$ is the most voted class for sample $x$ and $v_{c_1}$ the number of related votes, $c_2$ is the second most popular class and $v_{c_2}$ the number of corresponding votes [15]. This margin’s range is from 0 to +1.

$$\text{margin}(x) = \frac{v_{c_1} - v_{c_2}}{\sum_{c=1}^L(v_c)}$$

$$= \frac{\max_{c=1, \ldots, L}(v_c) - \max_{c=1, \ldots, L \neq c_1}(v_c)}{L}$$

(1)
where $T$ represents the number of base classifiers in the ensemble.

To efficiently handle classification problems, we concentrate on smaller margin instances which are generally closer to class boundaries and thus are potentially more informative [15, 16].

In all our experiments, bagging was used to create an ensemble involving Classification and Regression Tree (CART) [17] as base classifier. And all the reported results are mean values of a 10-time calculation.

3. ENSEMBLE TRAINING DATA MANIPULATION FOR REMOTE SENSING

Remote sensing data, especially in dense urban scenes, present three main problems: (1) there are many mislabeled instances in the data, (2) the data are highly imbalanced, (3) there is a large number of redundant instances. In [5], we have shown that both redundant and mislabeled instances are mostly present among high margin instances. In this section, we first construct an ensemble classifier at data level. A new bagging algorithm [5] is then applied to reduce training set size, minimize data redundancy, and improve the classification accuracy.

Bagging relies on bootstrap sampling over training data to produce diversity [7]. The diversity is derived from the differences between the training sets of base classifiers [18]. However, bootstrapped training sets become more and more similar as redundancy is increasing. Redundancy can significantly decrease diversity, thus degrading the performance of bagging.

3.1. Margin-based bagging

Generally, central instances occur in larger numbers than boundary instances. Consequently, central instances are more likely to be similar. Thus, there is potentially more redundancy in central instances. We apply here a new bagging method to map remote sensing data. This technique selects the most informative instances based on their margins [5]. It trains first the committee of bootstrapped base classifiers on the complete training set and then removes iteratively the data points on which the committee most agrees according to the unsupervised ensemble margin previously defined. This amounts to removing typical (highest margin) instances during the learning process.

3.2. Application to remote sensing data

3.2.1. Remote sensing data sets

Three remote sensing datasets (cf. table 1) were involved in the experiments. The two first datasets are an airborne urban image depicted on figure 1, as well as a multi-source data, resulting from the combination of Lidar data with this RGB image data. The third dataset is a Landsat satellite data from UCI Repository [19]. All datasets were randomly selected from the original data [5]. The ensemble size was 100. The size $M$ of the subset of instances to remove at each pruning step was set to 5% of the whole training set.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Training</th>
<th>Test</th>
<th>Attributes</th>
<th>Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airborne image</td>
<td>38764</td>
<td>156896</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Multi-source</td>
<td>38764</td>
<td>156896</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Satellite UCI</td>
<td>2000</td>
<td>2000</td>
<td>35</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 1. Remote sensing data sets

![Airborne urban image sample of size 485 × 640 pixels](image)

Fig. 1. Airborne urban image sample of size 485 × 640 pixels

3.2.2. Overall classification performance

Table 2 presents the average and standard deviation of the classification accuracy obtained on test set by the chosen ensemble that led to the maximum classification accuracy on training set, as well as the pruned training set size (percentage of lowest margin instances of the whole training set) of this optimal ensemble. The training data sizes of all three datasets were significantly reduced, especially in case of multi-source data for which 85% of the full training set was removed. The best increase in classification accuracy (2.5%) was achieved for dataset Satellite UCI.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Acc. W (%)</th>
<th>Acc. R (%)</th>
<th>Size R (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airborne image</td>
<td>81.85 ± 0.02</td>
<td>81.90 ± 0.20</td>
<td>18.0 ± 2.58</td>
</tr>
<tr>
<td>Multi-source</td>
<td>94.04 ± 0.01</td>
<td>94.62 ± 0.04</td>
<td>15.0 ± 0.00</td>
</tr>
<tr>
<td>Satellite UCI</td>
<td>86.95 ± 0.33</td>
<td>89.49 ± 1.76</td>
<td>57.8 ± 7.14</td>
</tr>
</tbody>
</table>

Table 2. Accuracy on test sets by ensembles with whole training set (W), reduced training set (R) and related reduced size.

3.2.3. Classification performance per class

Table 3 compares the classification accuracy achieved, on test set, for the most difficult class of each dataset by both our
bagging method and classic bagging with a full training set. This table shows that our method significantly outperforms classic bagging in terms of handling complex classes. The increase in accuracy on the most difficult class is over 8% for both Multi-source and Satellite UCI datasets. These results confirm that our method is suitable for the classification of imbalanced remote sensing data.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Accuracy W (%)</th>
<th>Accuracy R (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airborne image</td>
<td>41.88 ± 0.39</td>
<td>43.61 ± 0.65</td>
</tr>
<tr>
<td>Multi-source</td>
<td>59.96 ± 0.46</td>
<td>67.99 ± 0.36</td>
</tr>
<tr>
<td>Satellite UCI</td>
<td>46.47 ± 1.09</td>
<td>55.69 ± 1.51</td>
</tr>
</tbody>
</table>

Table 3. Classification accuracy for the most difficult class by ensembles with whole training set (W) and reduced training set (R), on test sets.

3.2.4. Selection of training data

Figure 2 exhibits the best training set (depicted in black) provided by our method with all of multi-source data as training set. The training data that have been discarded are depicted in white. The best training instances appear to be located mainly at class boundaries (eg. building facades) and in small classes (eg. vegetation). Indeed, they have been chosen as the lowest margins and hence are likely to belong to class boundaries or minority classes. Table 4 shows the confusion matrix for the classification of all the data by bagging trees using the best training instances selected by our bagging method. Table 5 exhibits the confusion matrix obtained using bagging trees with all training instances. The classification accuracy is again highly improved on minority classes using our bagging method. The best improvement in accuracy is about 12% (natural ground class).

Fig. 2. Best training instances in multi-source urban data (black pixels) selected by margin-based bagging. White pixels are removed. Gray regions are not labeled.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Predicted</th>
<th>Observed</th>
<th>Artificial</th>
<th>Ground</th>
<th>Build.</th>
<th>Natural</th>
<th>Ground</th>
<th>Vege.</th>
<th>Class error %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial Gr.</td>
<td>90664</td>
<td>2420</td>
<td>327</td>
<td>543</td>
<td>1</td>
<td>3.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Building</td>
<td>2682</td>
<td>85996</td>
<td>57</td>
<td>280</td>
<td>1</td>
<td>3.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natural Gr.</td>
<td>1201</td>
<td>58</td>
<td>3056</td>
<td>8</td>
<td>1</td>
<td>29.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vegetation</td>
<td>1580</td>
<td>619</td>
<td>73</td>
<td>6096</td>
<td>1</td>
<td>27.16</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Confusion matrix for multi-source data using a bagging classifier with best training set selected by margin-based bagging. Total error rate=5.04%.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Predicted</th>
<th>Observed</th>
<th>Artificial</th>
<th>Ground</th>
<th>Build.</th>
<th>Natural</th>
<th>Ground</th>
<th>Vege.</th>
<th>Class error %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial Gr.</td>
<td>90212</td>
<td>2591</td>
<td>330</td>
<td>821</td>
<td>1</td>
<td>3.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Building</td>
<td>2999</td>
<td>85658</td>
<td>25</td>
<td>333</td>
<td>1</td>
<td>3.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natural Gr.</td>
<td>1686</td>
<td>67</td>
<td>2557</td>
<td>13</td>
<td>1</td>
<td>40.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vegetation</td>
<td>1951</td>
<td>801</td>
<td>63</td>
<td>5553</td>
<td>1</td>
<td>33.65</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Confusion matrix for multi-source data using a bagging classifier with whole training set. Total error rate=5.94%.

4. ENSEMBLE PRUNING FOR REMOTE SENSING

In this section, we propose to construct an ensemble classifier at classifier level by reducing the number of base classifiers. In ensemble learning, this is called ensemble pruning or ensemble selection.

Ensemble pruning focuses on searching for a good subset of ensemble members that performs as well as or better than the original ensemble [20, 21]. Many ensemble pruning methods have been proposed in literature and can be classified into the following categories [22]: ordering-based, clustering-based, and optimization-based methods. The simplest techniques are ordering-based methods. In [23], it is shown that the latter are competitive, in terms of classification accuracy, with computationally more costly methods that directly select optimal sub-ensembles. Our ensemble pruning method [5] belongs to this category.

4.1. Margin-based ensemble pruning

Our ensemble pruning algorithm relies on an innovative margin-based criterion on choosing better classifiers [24, 5]. This measure involves once again the alternative definition of ensemble margin (equation 1) and is defined as:

\[ H(X) = -\frac{1}{|X|} \sum_{i=1}^{X} \log(\text{margin}(x_i)) \] (2)

where \( x_i \) is an instance of set \( X \).

This pruning method is one of overproduce – and – choose methods which can be divided into overproduction and choose phases [25]. Firstly, the overproduced classifiers, \( C_{t:t=1, \ldots, T} \), are used to classify a pruning set \( V \) and calculate the margin, defined in equation 1, of each instance.
in this set. Then, the margin-based criterion is computed for each classifier \( C_t \):

\[
H_t(V) = -\frac{1}{N} \sum_{i=1}^{N} \log(\text{margin}(x_i)) \quad \forall(x_i, y_i) \in V / C_t(x_i) = y_i
\]

(3)

where \( x_i \) is an instance of set \( V \) that has been well classified by classifier \( C_t \). Thus, \( 0 \leq H_t(V) \leq \log(T) \), with \( \text{margin}(x_i) \geq \frac{1}{T} \) for a given odd number \( T \) of base classifiers.

All classifiers are then ordered according to this ranking measure thus leading to a sorted list of classifiers with potentially a decreasing reliability. Finally, we choose the \( T^* \) first ordered base classifiers to compose a pruned ensemble which had the best overall accuracy for pruning set.

4.2. Application to remote sensing data

Our ensemble pruning method is compared to accuracy ordering-based ensemble pruning. The latter is the most similar competing method to our approach. Indeed, accuracy ordering pruning evaluates each single classifier on all instances. Our method focuses on single classifier’s performance but just on small margin instances. Our approach will also be compared to complete ensemble bagging, in which no pruning takes place.

4.2.1. Remote sensing data sets

The ensemble pruning methods were applied to the three available remote sensing datasets, as shown on table 6. Each dataset has been divided into three parts: training set, pruning set and test set. The pruning set is used to select the sub-ensemble that leads to the maximum classification accuracy. For all datasets, the size of the original ensemble is 500.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Margin ordering</th>
<th>Accuracy ordering</th>
<th>Complete bagging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airborne image</td>
<td>66.67</td>
<td>67.21</td>
<td>67.61</td>
</tr>
<tr>
<td>Multi-source</td>
<td>41.15</td>
<td>42.08</td>
<td>46.14</td>
</tr>
<tr>
<td>Satellite UCI</td>
<td>52.13</td>
<td>52.80</td>
<td>53.53</td>
</tr>
</tbody>
</table>

Table 7. Maximum classification error rate (average) per class by the selected sub-ensemble of both ordering-based pruning methods, and complete bagging, for all data sets.

5. CONCLUSION

In this work, we have shown that the involvement of low margin instances in the learning process is an appealing solution for building efficient ensemble classifiers for remote sensing data. We have considered two levels to construct reliable ensembles for mapping remote sensing data: data and classifier levels. At data level, we applied our margin-based bagging method. The obtained results show that it increases the classification accuracy, particularly in case of minority classes, and significantly reduces the training set size. At classifier level, we applied our margin-based ensemble pruning method. Our method not only significantly reduces the complexity of the original ensemble but also increases the classification accuracy for difficult and/or minority classes.
6. REFERENCES


