ENSEMBLE METHODS IN CLASSIFICATION OF REMOTE SENSING IMAGES

Lorenzo BRUZZONE

E-mail: lorenzo.bruzzone@ing.unitn.it
Web pages at: http://www.ing.unitn.it/~bruzzone/RSLab
Outline

- Introduction: remote sensing and classification problems;
- Why Multiple Classifier Systems (MCSs) in classification of remote sensing images?
- Applications of MCSs in remote sensing:
  - Supervised classification;
  - “Partially supervised” classification.
- MCS in “partially supervised” classification problems;
- Examples and discussion;
- Conclusions.
Introduction

What is remote sensing?
Example of remote sensing images acquired by a multispectral sensor and a multiband multipolarimetric synthetic aperture radar
Introduction

Classification of remote sensing images is one of the most complex applications of pattern recognition. Several factors make it very critical to perform classification with a high accuracy:

- atmospheric noise;
- soil moisture;
- sensor calibration problems;
- geometrical resolution (presence of “mixed pixels”);
- multisource/multisensor data;
- hyperspectral data (sensors of the last generation).
Introduction

Typical classification approaches used in remote sensing

- Standard statistical classifiers (maximum likelihood, Bayes classifier, k-nearest neighbor, etc.);
- Neural Networks (multilayer perceptrons, radial basis function neural networks, probabilistic neural networks, structured neural networks, etc.);
- Fuzzy classifiers;
- Knowledge-based approaches;
- Empirical approaches.
MCSs in Classification of Remote Sensing Images

**Why using MCSs to classify remote sensing data?**

- High complexity of the classification problems;
- Need to obtain reliable and accurate classification systems;
- Availability of several classification algorithms widely tested on remote sensing images.

However, although some papers have been published on this topic in the literature, **MCSs in remote sensing are under-illuminated** with respect to other methodological approaches.
Applications of MCSs in Classification of Remote Sensing Images

Two main problems have been addressed in the analysis of remote-sensing images by using MCSs:

- Supervised Classification;
- “Partially Supervised” Classification of Multitemporal Images.
MCSs in Supervised Classification

MCSs can be used in remote sensing to effectively address the **supervised classifications** of:

- Multisource/multisensor images;
- Multispectral (or single source) images;
- Hyperspectral images.
The ever increasing availability of remote-sensing images acquired by different sensors makes it mandatory to develop effective multisource/multisensor classification approaches to exploit the complementarities of data acquired by different sensors.

The most common approach applied to classification of multisource/multisensor data consists in using non-parametric classification algorithms (e.g. neural networks) according to the “stacked vector” method.

In 1992, Benediktsson et al. [1] introduced the use of MCSs to classify multisensor/multisource remote sensing images.
MCSs in Supervised Classification: Multisource/Multisensor Images

Classifier 1

Classifier 2

Combination Strategy

Land-cover map

S1
Multispectral image

S2
SAR image

SN
Ancillary data

Multisource/Multisensor Images

S1
Classifier 1

S2
Classifier 2

SN
Classifier N

International Summer School on Neural Nets E.R. Caianiello, Vietri sul Mare, September 2002
The most popular combination approach in multisource/multisensor remote sensing classification problems is the “Statistical Consensus Theory” [1-2].

Statistical Consensus Theory: hybrid classification approach based on consensus from different classifiers, each one specialized on a specific information source. Different strategies are usually adopted to obtain the consensus by considering the reliability of each information source:

- Linear Opinion Pool (LOP);
- Logarithmic Opinion Pool (LOGP);
- Neural networks.
MCSs in Supervised Classification: Multisource/Multisensor Images

Consensus rules [1-2]

- **Linear Opinion Pool (LOP)** ($\lambda_i$ is the reliability factor of the i-th information source):
  \[ C_j(X) = \sum_{i=1}^{N} \lambda_i P(\omega_j | x_i) \]

- **Logarithmic Opinion Pool (LOGP):**
  \[ \log(L_j(X)) = \sum_{i=1}^{N} \lambda_i \log [P(\omega_j | x_i)] \]

- **Neural Networks**: multilayer perceptrons neural networks are used to integrate classifiers.
Other multiple classifier approaches investigated in the context of remote sensing are [5]:

- Bagging;
- Boosting.

Depending on the specific problem (e.g. sensors used, geographical area considered, acquisition conditions, etc.), the different combination strategies result in different effectiveness.
MCSs in Supervised Classification: Multispectral Images

As in other application domains, also in remote sensing ensemble methods have been used in classification problems characterized by a single information source (e.g. classification of multispectral images) to increase both the reliability and accuracy of the classification process.
Different ensemble methods have been investigated with results that depend on the specific data set considered:

- **Combination-based approaches [7, 10]**
  - Majority Voting;
  - Bayesian Average;
  - Belief Functions;
  - Consensus Theory;
  - Consensus-based voting and rejection scheme.

- **Dynamic classifier selection** by using measures of local accuracy [8-9].
MCSs in Supervised Classification: Hyperspectral Images

- **Hyperspectral images**: images acquired by **hyperspectral sensors** in 100-300 spectral channels (usually in the spectral range between visible and infrared).

- **Example**: AVIRIS sensor, 224 spectral channels, spectral resolution 10 [nm], radiance quantized on 12 bit.

- **Advantages**: very detailed analysis of the spectral signature of land-covers (intrinsic capability to discriminate among several land-cover classes).
MCSs in Supervised Classification: Hyperspectral Images

- **Disadvantages:** Hughes phenomenon.

- **Typical solution to the Hughes phenomenon problem:** application of a feature extraction/selection process to hyperspectral images.

- **Alternative solution:** adoption of MCSs [10-12].
MCSs in Supervised Classification: Hyperspectral Images

How exploiting MCSs to classify hyperspectral images?

**Basic idea:** define an ensemble of classifiers in which every classification algorithm is applied to a different subset of spectral channels. The combination of the results obtained by such classifiers produces the final classification map.

According to this procedure, each classifier works in a reduced feature space. This makes it is possible to overcome the problems involved from the small number of training samples versus the total dimensionality of the feature space.
Theoretical Problem: how selecting subsets of spectral channels?

Different solutions have been proposed in the literature:

- features are divided in subsets composed of spectral channels close each others in the spectral domain;
- features can be divided in different subsets so that the resulting subsets are as uncorrelated as possible.

The obtained subsets are then modeled as independent data sources (“equivalent sources”) and standard combination approaches (e.g. LOGP) are applied.
MCSs in “Partially Supervised” Classification

- **Problem addressed:** periodical monitoring of extended geographical areas by classifying multitemporal images (environmental risk, updating of land-cover maps in GIS, etc.)

- **Constraint:** ground truth information is not available for all the images to be classified (ground truth information is only available for one image in the considered temporal sequence of data).

- **Purpose:** develop robust and reliable classification systems that can analyze accurately also images of the considered site for which ground truth is not available.
MCSs in “Partially Supervised” Classification

Solution proposed in the literature

- Develop partially supervised classification procedures [16].
- Exploit methodologies for the combination of classifiers to increase the accuracy and reliability of each partially supervised classifier [13-15].

Two main approaches have been proposed:

- MCSs based on single date partially supervised classifiers [14, 16];
- MCSs based on multitemporal partially supervised classifiers [13, 15, 17].
MCSs in “Partially Supervised” Classification: Single Date Classifiers

$X_1 \xrightarrow{\text{Image at } t_1} \text{Classifier} \xrightarrow{\text{Supervised estimation of classifier parameters}} \text{Land-cover map at } t_1$ 

$\text{Training set } Y_1$

$\text{Estimated from the training set } Y_1$

$\text{TIME } t_1$

$X_j^l \in \omega_i \Leftrightarrow \omega_i = \arg \max_{\omega_k \in \Omega} \{\hat{P}_1(\omega_k) \hat{P}_1(X / \omega_k)\}$
MCSs in “Partially Supervised” Classification: Single date Classifiers

\[ \begin{align*}
\text{Image at } t_2 & \quad \rightarrow \quad \text{Classifier} \\
\text{Training set } Y_2 & \quad \rightarrow \quad \text{Partially supervised estimation of classifier parameters} \\
\text{Values of the classifier parameters estimated in a supervised way at } t_1 & \quad \rightarrow \quad \text{Land-cover map at } t_2
\end{align*} \]

\[ x_j^2 \in \omega_i \iff \omega_i = \arg \max_{\omega_k \in \Omega} \left\{ \hat{p}_2(\omega_k) \hat{p}_2(X / \omega_k) \right\} \]

\text{Partially supervised estimation}
Let $X_1$ and $X_2$ be two remote-sensing images acquired at two times $t_1$ and $t_2$. Let us assume that only a training set $Y_1$ related to $X_1$ is available.

Let $\Omega=\{\omega_1, \omega_2, \ldots, \omega_C\}$ be the set of land-cover classes that characterizes the study area at both acquisition dates. Let $x_j$ be the feature vector associated with the j-th pixel of $X_i$.

Let us consider a classifier based on the Bayes rule for minimum error:

\[
\text{TIME } t_1 \quad x_j^1 \in \omega_i \iff \omega_i = \arg\max_{\omega_k \in \Omega} \left\{ \hat{P}_1(\omega_k) \hat{P}_1(X / \omega_k, \theta_1) \right\} \\
\text{TIME } t_2 \quad x_j^2 \in \omega_i \iff \omega_i = \arg\max_{\omega_k \in \Omega} \left\{ \hat{P}_2(\omega_k) \hat{P}_2(X / \omega_k, \theta_2) \right\}
\]

Estimated from the training set $Y_1$
MCSs in “Partially Supervised” Classification: Single date Classifiers

Development of each partially supervised classifier: the parameters estimated from the supervised training at the time $t_1$ can be updated at $t_2$ by using the information associated with the distribution of the second image $p_2(x)$.
MCSs in “Partially Supervised” Classification: Single date Classifiers

The distribution of the $X_2$ image can be described as a mixture density composed of as many components as classes to be recognized:

$$p_2(X) = \sum_{i=1}^{C} P_2(\omega_i)p_2(X / \omega_i, \theta_2^i)$$

The density functions of classes $p_2(X / \omega_i, \theta_2^i)$ depend on the parameter vector $\theta_2^i$ of the considered classifier. The number and type of components of the vector depend on the type of classifier considered.

The computation of $P_2(\omega_i)$ and $\theta_2^i$ becomes an unsupervised mixture estimation problem, which can be solved by applying the iterative expectation-maximization (EM) algorithm [14,16].
MCSs in “Partially Supervised” Classification: Single date Classifiers

Given the complexity inherent with the partially supervised classification, each classifier is intrinsically less reliable and accurate than the corresponding supervised one, especially in complex data sets.

The exploitation of both ensembles of partially supervised classifiers and suitable combination strategies may increase the reliability and (possibly) the accuracy of the classification system.
MCSs in “Partially Supervised” Classification: Single date Classifiers

Images at $t_1$

Classifier 1

Classifier N

Multiple classifier approach

Supervised estimation of the classifier parameters

Training set $Y_1$

Land-cover map at $t_1$
MCSs in “Partially Supervised” Classification: Single date Classifiers

Images at $t_2$

Classifier 1

Classifier N

Partially supervised estimation of the classifier parameters

Values of the classifier parameters at $t_1$

Multiple classifier approach

Land-cover map at $t_2$
How to define partially supervised classification algorithms to be included in the ensemble?

Two main approaches have been proposed to derive suitable classification algorithms:

- maximum likelihood classifier [16];
- radial basis function (RBF) neural networks [14].

**Combination strategies**

- **Constraints**: the classifier combination should be performed according to unsupervised strategies (no training set is available at \( t_2 \)).
- **Adopted strategies**: majority voting; Bayesian combination.
MCSs in “Partially Supervised” Classification: Multidate Classifiers

How to improve the performances of a partially supervised MCS?

To improve the effectiveness of the system, multidate classifiers can be used (instead of single date classifiers) in the considered ensemble of algorithms [13,15, 17]. In particular:

- each member of the ensemble of classifiers can be defined in the framework of the cascade classification decision rule (exploitation of temporal correlation between images);

- simple strategies for the ensemble design can be defined by taking into account the peculiarities of the cascade classification approach.

Expected improvements: the exploitation of temporal correlation between images may further increase both the classification accuracy at time $t_2$ and the reliability of the classification system.
Partially supervised classifiers are developed in the framework of the cascade classification decision rule:

\[ x_j^2 \in \omega_m \quad \text{if} \quad P(\omega_m / x_j^1, x_j^2) = \max_{\omega_k \in \Omega} \{ P(\omega_k / x_j^1, x_j^2) \} \]

Under the assumption of conditional independence in the time domain, the rule can be rewritten as:

\[ x_j^2 \in \omega_m \quad \text{if} \quad \sum_{n=1}^{C} p(x_j^1 / \omega_n) p(x_j^2 / \omega_n) P(\omega_n, \omega_m) = \max_{\omega_k \in \Omega} \left\{ \sum_{n=1}^{C} p(x_j^1 / \omega_n) p(x_j^2 / \omega_n) P(\omega_n, \omega_k) \right\} \]

Estimated on the training set \( Y_1 \)
MCSs in “Partially Supervised” Classification: Multidate Classifiers

Partially supervised estimation of classifier parameters

Training set $Y_1$

Training set $Y_2$

Prior information on $P(\omega_n, \omega_k)$ (if available)

Land-cover map at $t_2$

$p(x_j^1/\omega_n)$

$p(x_j^2/\omega_k)$

$p(x_1, X_2)$

$p(x_j^2/\omega_k)$

$p(\omega_n, \omega_k)$
MCSs in “Partially Supervised” Classification: Multidate Classifiers

Development of each partially supervised multidate classifier

The parameters of each partially unsupervised classifier can be estimated by: i) exploiting the information contained in the training set \( Y_1 \); ii) using the information associated with the joint density \( p(x_1, x_2) \).
MCSs in “Partially Supervised” Classification: Multidate Classifiers

The \( p(x_1, x_2) \) can be described as a mixture density composed of as many components as the possible pairs of classes at the two dates:

\[
p(x_1, x_2) = \sum_{i=1}^{C} \sum_{j=1}^{C} p(\omega_j | \omega_i \theta_u^i) p(x_1 | \omega_j \theta_u^i) p(x_2 | \omega_i \theta_u^i)
\]

The computation of the parameter vector \( \theta_u^i \) becomes a partially unsupervised mixture estimation problem, which can be solved by applying the iterative expectation-maximization (EM) algorithm [13].

The number and type of components of the vector \( \theta_u^i \) depend on the type of classifier considered.
MCSs in “Partially Supervised” Classification: Multidate Classifiers

\[
\theta^1 = [\theta^1_s, \theta^1_u] \quad \theta^N = [\theta^N_s, \theta^N_u]
\]

Partially supervised estimation of the cascade classifier parameters

Image at \( t_1 \)

Image at \( t_2 \)

Training set \( Y_1 \)

Training set \( Y_2 \)

Land-cover map at \( t_2 \)
MCSs in “Partially Supervised” Classification: Multidate Classifiers

- **Problem**: Design an ensemble of multitemporal classifiers based on the cascade classification decision rule.

- **Proposed solution**: two cascade classifiers (which exploit different techniques to estimate density functions of classes, and consequently prior joint probabilities of classes) have been developed. The estimation techniques selected are based on [13]:

  - Maximum likelihood classifier;
  - Radial basis function (RBF) neural networks.
MCSs in “Partially Supervised” Classification: Multidate Classifiers

Definition of ensembles of multidate classifiers: ensembles of cascade classifiers are generated by defining hybrid classifiers obtained by exchanging the estimates of the prior joint probabilities of classes performed by the previously described ML and RBF cascade classifiers [13]:
MCSs in “Partially Supervised” Classification: Multidate Classifiers

Image at $t_1$

$X_1$

Image at $t_2$

$X_2$

ML Classifier

ML-Hybrid Classifier

RBF-Hybrid Classifier

RBF Classifier

Partially unsupervised estimation of the classifier parameters

$\theta_{ML} = [\theta_{s}^{ML}, \theta_{u}^{ML}]$

$\theta_{RBF} = [\theta_{s}^{RBF}, \theta_{u}^{RBF}]$

Multitemporal combination strategy

Land-cover map at $t_2$

Training set $Y_2$

Training set $Y_1$
MCSs in “Partially Supervised” Classification: Multidate Classifiers

- How to combine cascade classifiers?
  Standard combination strategies (which are modified to consider the multitemporal feature of cascade classifiers) can be considered.

- Combination strategies adopted
  - Multitemporal Majority Voting;
  - Multitemporal Bayesian Average.
MCSs in “Partially Supervised” Classification: Example

- **Considered test site:** area of Lake Mulargias (Sardinia, Italy)
- **Considered data:** 2 Thematic Mapper images acquired in September 1995 ($t_1$) and July 1996 ($t_2$).

![September 1995](image1.jpg) ![July 1996](image2.jpg)
### MCSs in “Partially Supervised” Classification: Example

<table>
<thead>
<tr>
<th>Land-cover classes</th>
<th>Number of patterns (September 1995)</th>
<th>Number of patterns (July 1996)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training set</td>
<td>Test set</td>
</tr>
<tr>
<td>Pasture</td>
<td>530</td>
<td>605</td>
</tr>
<tr>
<td>Forest</td>
<td>101</td>
<td>100</td>
</tr>
<tr>
<td>Urban area</td>
<td>163</td>
<td>169</td>
</tr>
<tr>
<td>Water body</td>
<td>517</td>
<td>533</td>
</tr>
<tr>
<td>Vineyard</td>
<td>249</td>
<td>235</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1560</strong></td>
<td><strong>1642</strong></td>
</tr>
</tbody>
</table>

Not used with the partially supervised approach
MCSs in “Partially Supervised” Classification: Example

Considered partially supervised cascade classifiers

- 1 maximum likelihood cascade classifier (Gaussian assumption);
- 1 RBF neural-network cascade classifier (with 50 hidden neurons);
- 2 “hybrid” cascade classifiers derived from the previous ones.

To exploit the non-parametric nature of the RBF neural classifier, 5 texture features (based on the Gray-Level Co-occurrence matrix) were given as input to the RBF and RBF-hybrid classifiers, in addition to the 6 TM channels.
MCSs in “Partially Supervised” Classification: Example

Design of experiments

- **Experiment 1**: analysis of the effectiveness of each partially supervised cascade classifier and of the resulting ensemble of partially supervised multidate classifiers.

- **Experiment 2**: analysis of the effectiveness of the partially supervised MCS when the failure of a partially supervised classifier is simulated.

- **Experiment 3**: comparisons with the effectiveness of a standard classifier trained on the $t_1$ image and applied to the $t_2$ image.
MCSs in “Partially Supervised” Classification: Example

Overall classification accuracies exhibited on the July 1996 test set \(t_2\) image by the partially supervised multidate classifiers and the multiple classifier architecture.

<table>
<thead>
<tr>
<th>Classification Technique</th>
<th>Overall Classification Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Likelihood</td>
<td>91.48</td>
</tr>
<tr>
<td>RBF</td>
<td>96.10</td>
</tr>
<tr>
<td>ML-hybrid</td>
<td>91.79</td>
</tr>
<tr>
<td>RBF-hybrid</td>
<td>95.38</td>
</tr>
<tr>
<td>Combination: Majority Voting</td>
<td>96.56</td>
</tr>
<tr>
<td>Combination: Bayesian Average</td>
<td>94.77</td>
</tr>
</tbody>
</table>
MCSs in “Partially Supervised” Classification: Example

Overall classification accuracies exhibited on the $t_2$ test set by the partially supervised multidate classifiers and the multiple classifier architecture (case in which the failure of the partially supervised training of the RBF cascade classifier is simulated).

<table>
<thead>
<tr>
<th>Classification Technique</th>
<th>Overall Classification Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Likelihood</td>
<td>91.48</td>
</tr>
<tr>
<td>RBF</td>
<td>67.68</td>
</tr>
<tr>
<td>ML-hybrid</td>
<td>91.74</td>
</tr>
<tr>
<td>RBF-hybrid</td>
<td>72.75</td>
</tr>
<tr>
<td>Combination: Majority Voting</td>
<td>95.90</td>
</tr>
<tr>
<td>Combination: Bayesian Average</td>
<td>92.46</td>
</tr>
</tbody>
</table>
MCSs in “Partially Supervised” Classification: Example

Overall classification accuracies exhibited on the $t_2$ test set by standard single date supervised ML and RBF classifiers trained on the $t_1$ image.

<table>
<thead>
<tr>
<th>Classification Technique</th>
<th>Overall Classification Accuracy (%) at $t_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Likelihood</td>
<td>35.91</td>
</tr>
<tr>
<td>RBF</td>
<td>72.85</td>
</tr>
</tbody>
</table>
MCSs in “Partially Supervised” Classification: Example
Conclusions

Classification of remote-sensing data is a complex problem, given the several critical factors that may affect this process.

Lately, the remote sensing community is realizing that MCSs are a very promising approach for several remote sensing applications.

MCSs can be effectively used in supervised classification of multispectral images, in multisource/multisensor problems, and in the analysis of hyperspectral data.

MCSs seem also an interesting approach to address complex and critical classification problems (e.g. “partially supervised” classification problems), which are very relevant from the application viewpoint.
Conclusions

MCSs in Remote Sensing: Open Issues

- How to select classification approaches to be included in the ensemble of classifiers? How to define the architecture of each classifier?

- How to identify the most suitable combination strategy to be used for solving a classification problem? Combination or selection strategies?

- What kinds of remote sensing problems can really benefit from the use of MCSs?

- How to define subset of hyperspectral images to be given as input to different classifiers?

- How to implement MCSs in standard GIS software?
Conclusions

MCSs in Remote Sensing: Advantages

- In remote sensing, the effectiveness of different classification algorithms is strongly data-dependent (in some cases, data-dependent on the same geographical area). Consequently, the use of different classification algorithms integrated in an ensemble can increase the reliability of a classification system.

- End-users require ever increasing accuracies. MCSs represent a convincing approach to increase the performances of automatic classification systems.
MCSs in Remote Sensing: Advantages

• New generation satellites acquire different types of data that contain complementary information. Extracting the information from the data requires to design a specific classification procedure for each specific sensor. MCSs can integrate the results provided by different classifiers.

• MCSs make it possible to perform classification of hyperspectral data by exploiting all the available spectral channels.
Conclusions

MCSs in Remote Sensing: Disadvantages

- Difficult to introduce in the application domain (e.g. in end-users oriented GIS system) a complex classification approach that requires different choices: classification algorithms, architectures, combination/selection strategies, etc.

- End-users are still reluctant to introduce advanced automatic classification approaches in the analysis of data (low comprehension of the methods, too many parameters to tune, etc.). This may be very critical with MCSs.

- The remote-sensing community has not defined “standardized procedures” for using MCSs in real applications.
MCS in Supervised Classification: Multisource/Multisensor Images


MCS in Supervised Classification: Multisource/Multisensor Images


References

MCS in Supervised Classification: Multispectral Images


References

MCS in Supervised Classification: Hyperspectral Images


MCS in “Partially Supervised” Classification Problems


MCS in “Partially Supervised” Classification Problems


