An Application of Combined Classifiers to Remotely Sensed Images

SANTOS, R. V.[1], VELLASCO, M.B.R.[1,2], FEITOSA, R. Q.[1,2] SIMÕES, M. [2], TANSCHEIT, R.[1]

[1]Dept. of Electrical Engineering ;
Catholic University of Rio de Janeiro
Rua Marquês de São Vicente, 225
22453-900, Rio de Janeiro
BRAZIL

[2]Dept. of Computer Engineering
State University of Rio de Janeiro
Rua São Francisco Xavier, 524
20550-013, Rio de Janeiro
BRAZIL

Abstract: Studies in the area of Pattern Recognition have indicated that a classification model performs differently from one pattern class to another. This observation gave birth to the idea of combining the individual results of different classifiers to derive a consensus decision. This work investigates the potential of combining classifiers to remotely sensed images. A classifier system is built integrating the results of a feed-forward neural network and of a maximal likelihood classifier. Fuzzy Integrals are used as the combining strategy. Experiments carried out to evaluate the system, using a satellite image of an area undergoing a rapid degradation process, has shown that the combination may considerably improve the classification performance of the individual classifiers.

Key-Words: Combining Classifiers, Remote Sensing, Pattern Recognition, Fuzzy Integrals, Back Propagation, Maximal Likelihood

1 Introduction

The interest for images collected by digital multi-spectral imaging systems has been following the growing worldwide concern for environmental issues. Remotely sensed image data has been used for various Earth-science applications, such as mapping land use, geology, forest types, among others. A particular important application of this technology is the monitoring of the process of environmental degradation and the evaluation of the impact of preservationist measures.

Many classification methodologies have been applied to remotely sensed images [1][2][3], with the aim to achieve the best possible classification performance.

Studies in the area of Pattern Recognition have indicated that a classification model performs differently from one pattern class to another. This observation gave birth to the idea of combining the individual results of different classifiers to derive a consensus decision. Various classifier combination approaches have been proposed [4]. These studies have demonstrated that the combination may outperform each individual classifier.

The work reported in this paper has the objective to evaluate the potential of combining classifiers for land use classification of remotely sensed images.

A classification scheme is presented, where the results of two classifiers - a statistical maximal likelihood classifier and a multi layer perceptron neural network - are combined. The concept of fuzzy integrals are used as classifier combination strategy.

The experiments carried out on a satellite image of a region in Brazil under severe environmental degradation process have demonstrated that the concept of classifier ensemble may yield a better performance than single classifiers in the task of land use/ land cover classification.

This paper is organized as follows. The next section presents the theoretical concepts of fuzzy measures and fuzzy integrals. Section 3 describes the combined classifier scheme used in this work. Section 4 describes the evaluation experiments and section 5 discusses their results.

2 Theoretical Concepts

Fuzzy Integrals are functions that can be particularly useful for information fusion problems. They combine *evidences* to form a *hypothesis*, taking into account *expectations* about each evidence's relevance [5]. In mathematical terms, fuzzy integrals are non linear operations based on *fuzzy measures*, which are the generalization of classical measures. In the following subsections we provide a brief description of the theory of fuzzy measures and fuzzy integrals, focusing on how these concepts can be applied to combined classifiers.

2.1 Fuzzy Measures

A fuzzy measure is defined by a function that assigns a value in the [0,1] interval to each crisp set of the universal set [6]. Bringing this concept to the context of classifier combination, the fuzzy measure expresses the level of competence of a classifier in assigning a pattern to a class. It must be noted that this value is different from the concept of membership grade. In the later case a value is assigned by a classifier to a pattern, meaning its degree of membership to a particular class. The fuzzy measure, on the other hand, denotes the level of trust on this classifier, when evaluating the membership degree to a given class.

Formally, fuzzy measure is a function $g_{A\subseteq\Omega}$: $X\rightarrow[0,1]$, where Ω is the universal set comprising all crisp sets of a specific variable x.

A fuzzy measure is similar to a probability measure, except that it does not follow the *addition rule*, that is: if g is a fuzzy measure defined over a set Ω and A, $B \subset \Omega$ so that $A \cap B = \emptyset$, the equation $g_k(x_i \cup x_j) = g_k(x_i) + g_k(x_j)$ does not apply.

2.2 Applying Fuzzy Integrals to Combine Classifiers

Using the fuzzy measures definitions [6], Sugeno [7] defined the concept of fuzzy integrals. A fuzzy integral is a non-linear operation defined over measurable sets.

Let A be an object (pattern) to be classified. Let $T = \{t_1, t_2, \ldots, t_n\}$ be the set of possible classes to be chosen and $\mathbf{X} = \{x_1, x_2, \ldots, x_m\}$ is the set of available classifiers.

To each classifier to be combined, one must set fuzzy measures $g_k(x_i)$, denoting the competence of classifier x_i in the recognition of patterns belonging to class t_k. These densities may be set by experts or by training sets analysis. In this paper we consider $g_k(x_i)$ as the hit ratio at training phase for classifier x_i with respect to class t_k.

Let h_k: $\mathbf{X} \rightarrow [0,1]$ be a function which expresses how well the pattern fits in the class t_k according to the classifier $x_i \in \mathbf{X}$. If the cardinality of $\mathbf{X}$ is m, then $\mathbf{X}$ is arranged as $\{x_1, x_2, \ldots, x_m\}$ so that $h_k(x_1) \geq h_k(x_2) \geq \ldots \geq h_k(x_m) \geq 0$.

An ascending sequence of classifiers $\mathbf{Y}=\{y_1, y_2, \ldots, y_n\}$ will then be created, so that $y_1 = x_1$ and $y_i = y_{i-1} \cup x_i$, for $1 < i \leq n$, whereby the symbol $y_{i-1} \cup x_i$ denotes the classifier resulting from the combination of classifier y_{i-1} with classifier x_i.

Since the fuzzy measures do not follow the *addition rule*, Sugeno's proposal is applied to calculate the fuzzy measures for the new sequence of classifiers, according to the equation 1:

$$g_k(y_i) = g_k(y_{i-1} \cup x_i) = g_k(y_{i-1}) + g_k(x_i) + \lambda\, g_k(y_{i-1})\, g_k(x_i), \quad \textbf{(1)}$$

with $\lambda > -1$. The value of λ is always taken from the boundary condition $g(y_m)=1$, which means that the fuzzy measure of the classifier resulting from the combination of all original classifiers will be equal to 1. To determine λ, a n-1 degree equation must be solved:

$$\prod_{i=1}^{n} \left[1 + l\, g_k(x_i)\right] = 1 + l,\ l \neq 0 \quad \textbf{(2)}$$

Sugeno has proved that there is always an unique non-zero $\lambda \in (-1, \infty)$ that satisfies Eq.2.

The fuzzy integral (e_k) of the function h_k over $\mathbf{Y}$ with respect to g_k is given by Eq. 3 [5]:

$$e_k = \int_{\mathbf{Y}} h_k \circ g_k = \max_{i=1}^{m} \left[\min\left(h_k(x_i), g_k(y_i)\right)\right] \quad \textbf{(3)}$$

This expression is to be computed in two steps:

1. Obtain the minimum (or t-norm) between $h_k(x_i)$ and $g_k(y_i)$, for $1 \leq i \leq m$ and
2. Determine the maximum (or t-conorm) of the resulting sequence from phase 1.

There are several interpretations of fuzzy integrals. In this case, it is useful to understand them as a method of obtaining the maximum grade of agreement between competence $g_k(y_i)$ and confidence $h_k(x_i)$.

According to this procedure, a pattern will be assigned to the class having the highest value returned by the fuzzy integral.

So the complete algorithm for classifiers fusion, adapted from [5], is shown below in an informal way, presenting a clear view about a real world application of the previously seen concepts.

BEGIN classifusion,
FOR each class t_k
FOR each classifier x_i
determine $g_k(x_i)$ **END_FOR**
compute λ_k **END_FOR**
FOR each object **A**
FOR each class t_k
FOR each classifier x_i
read $h_k(x_i)$
END_FOR
compute the integral e_k
END_FOR
END_FOR
END
The t_k class with greatest integral value is chosen for the object A.

3 System Description

3.1 General Description

As shown in Fig.1, the proposed system uses two different classifiers to label a sample Landsat image and then proceeds to an information fusion stage which provides a definitive classification answer.

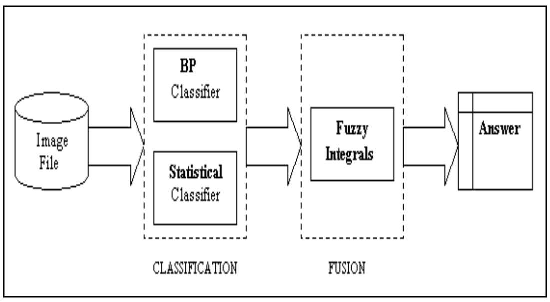

Fig. 1: Stages of the information fusion system.

The classification module is composed of a neural network Back-Propagation classifier and a Maximum Likelihood statistical classifier to produce the inputs for the fusion stage.

The values of h_k for the neural network and for the statistical classifier were taken, respectively, as the direct output of the network and as the *a posteriori* probability.

The heart of the fusion stage is a fuzzy integral algorithm, which works upon each individual classifier output. The theoretical foundations about fuzzy integrals were summarized in previous sections.

3.2 The Neural Network Classifier

In all the classification experiments performed in this work, the neural network had a *feed-forward* architecture with a single hidden layer. The learning algorithm, as mentioned before, was the *Back-Propagation*, with adaptive learning rate and fixed momentum [8].

The pixels of the image used in the experiments were defined by three 8-bits values, corresponding to the channels 3, 4 and 5 of the Landsat satellite images. So, each pixel was represented by 24 bits, 8 bits for each channel. This led us to use 24 inputs in the network.

The output layer was composed of 9 processors, one for each class of images.

The activation function used in both layers was the *log-sigmoid*, which held outputs always between 0 and 1.

Finally, two parameters were varied during the experiments performed here: first the number of processors number in the hidden layer; second, the number of points per class to be considered at the training phase.

3.3 The Statistical Classifier

The Statistical Classifier used here performs a Maximum Likelihood [9] pixel labeling. This technique is widely used in the image processing field [1], and therefore will not be further explained. Just to mention, the number of points per class in the training set was the only *unknown* parameter during the set of experiments carried out.

4 Experiments

4.1 Sample Image

All the experiments carried out were based on a RGB-mapped Landsat image, depicting a Brazilian micro-bay named “Agua-Limpa”. The image has 400 by 400 pixels, resulting in a total of 160.000 patterns to be classified (Fig.2).

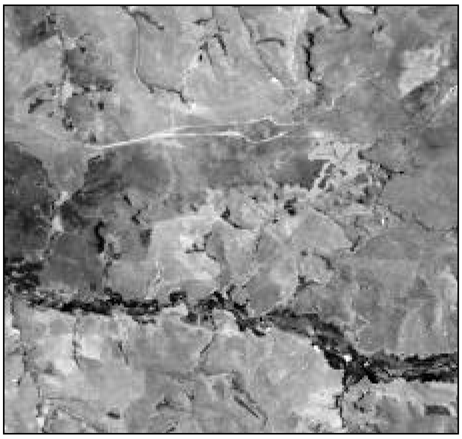

Fig. 2: “Agua-Limpa” micro-bay.

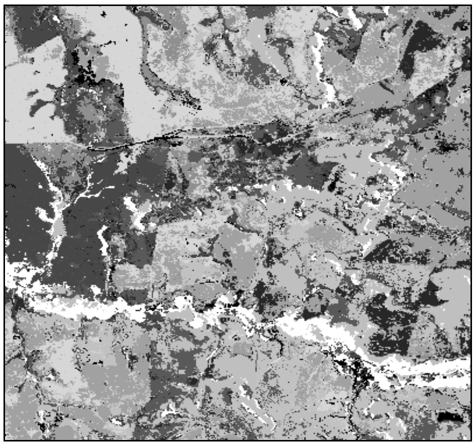

Fig. 3: Thematic map used as reference.

Classifier	Error Rate per Class									
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Average
Statistical	4,01	5,60	22,28	6,55	4,17	5,59	6,96	3,79	27,80	9,64
Neural Net	4,62	1,8	0	34,09	13,74	8,03	17,44	34,65	5,11	13,28

Table 1:: Neural Network and Maximum Likelihood performances

Classifier	Error Rate per Class									
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Average
Statistical	5,00	2,00	0,00	30,40	14,60	7,80	18,00	33,20	5,80	**12,98**
Neural Net	2,20	1,40	21,60	3,20	1,20	4,00	4,40	1,80	23,80	**7,07**

Table 2:Training performances for the individual classifiers

Classifier	Error Rate per Class									
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Average
Combined	2,55	2,33	0,68	12,67	6,13	4,81	7,78	8,44	4,76	**5,57**

Table 3: Final Classification Performance

To perform the supervised training it was necessary to have a reference classification. This reference was produced by an expert, mainly using his own knowledge along with some specific GIS tool (Fig.3).

5 Results

Before proceeding to the classifiers combination, we had to check out the individual results for each classifier. Using an error per class metric, that is, counting the percentage of pixels misclassified, the following performances have been achieved (Tab.1).

The data in the table concerning the neural network corresponds to an architecture with 12 neurons in the hidden layer. A total of 500 pixels per class were used in the training phase for both classifiers. Before this was done, we checked the training performance of these methods, to determine each one’s class “competence”, that is, each one’s *fuzzy* densities (Tab. 2).

Having set the training results as the fuzzy densities for the combination stage of our system (100 minus the training error rate), the execution of the fuzzy integral labeling procedure came up with the following (and final) classification response (Tab. 3):

In addition to the better overall performance achieved by the combination, there are some interesting remarks to be made.

First of all, one should notice the “smoother “ error distribution among classes in the final result, or the smaller standard deviation acquired, what can be seen as a kind of system robustness. Second, if we just look locally, there should be seen that for some classes (like the third) there is a great distortion between the classifiers results, distortion that is eliminated in the final mixture. So, one can have the feeling that combining classifiers can lead to a kind of “tuning” of some classes, perhaps some classes that need , more than others, to have better performance.

6 Final Comments

The potential of combining classifier to improve the classification accuracy of remotely sensed images has been investigated. A classification system was proposed, which combines the results of a statistical classifier and a feed-forward neural network. Fuzzy integrals were used as combination strategy.

The system was evaluated on a satellite image of an area suffering under a severe environmental degradation process. In the experiments for performance evaluation the combination attained an average performance considerably higher than the individual classifiers.

The experiments have also shown that the combination tend to equalize the performance among all classes, while improving the overall recognition rate.

These results encourage a further investigation of combined classifiers for this kind of application, with the aim to get a deeper understanding about what are the best combination strategies and in what circumstances.

References

[1] J. A. Richards, X. Jia - *Remote Sensing Digital Image Analysis – An Introduction*, 3rd Ed., Springer Verlag (1999).

[2] P. M. Mather, - *Computer Processing of Remotely-Sensed Images: An Introduction,* 2nd Ed, Willey (1999).

[3] T. M. Lilesand, R. W. Kiefer, *Remote Sensign and Image Interpretation,* 4th Ed., Wiley, (2000).

[4] J. Kitter, M. Hatef, R. P. W. Duin, J. Matas, *On Combining Classifiers,* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, NO. 3, pp. 226-239 March (1998).

[5] H. Tahani, J. M. Keller - *Information Fusion in Computer Vision Using the Fuzzy Integral*, IEEE Transactions On Systems, Man and Cybernetics, 20:733-741(1990).

[6] G. Klir, T. Folger, - *Fuzzy Sets, Uncertainty and Information*, Prentice-Hall International, Inc (1988).

[7] M. Sugeno - *Fuzzy measures and fuzzy integrals: A survey*, Fuzzy Automata and Decision Processes. Amsterdam: North Holland, pp.89-102 (1977)

[8] – S. Haykin - *Neural Networks – A Comprehensive Foundation*, Mcmillan College Publishing Co (1998).

[9] - R. Johnson, D. W. Wichern - *Multivariate Statistical Analysis*, 4th Ed. Prentice Hall (1998).