An Application of Combined Neural Networks to Remotely Sensed Images

Santos, R. V.[1], Vellasco, M. 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

e-mail: rafael@ele.puc-rio.br; marley@ele.puc-rio.br; raul@ele.puc-rio.br, maggie@eng.uerj.br, ricardo@ele.puc-rio.br;

ABSTRACT

Studies in the area of Pattern Recognition have indicated that in most cases a classifier performs differently from one pattern class to another. This observation gave birth to the idea of combining the individual results from different classifiers to derive a consensus decision. This work investigates the potential of combining neural networks to remotely sensed images. A classifier system is built by integrating the results of a plurality of feed-forward neural networks, each of them designed to have the best performance for one class. 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, have shown that the combination may yield a better performance than that of a single neural network.

Keywords: Combining Classifiers, Remote Sensing, Pattern Recognition

1. INTRODUCTION

The growing worldwide concern for environmental issues has increased the interest in images collected by digital multi-spectral imaging systems. Remotely sensed image data have been used for various Earth-science applications, such as mapping land use, geology, forest types, among others. A particularly 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 [Rich99] [Math99] [Lill00], with the aim of achieving 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 [Kitt98]; these studies have demonstrated that the combination may outperform each individual classifier.

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

A classification scheme is presented, that searchs for the best neural network for each class – the expert network. The results provided by these networks are then combined by using the concept of fuzzy integrals.

The experiments carried out on a satellite image of a region in Brazil under a severe environmental degradation process have demonstrated that the concept of classifier ensemble may yield a better performance than that of a single classifier 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 design procedure for expert neural networks; 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 the relevance of each evidence [Taha90]. In mathematical terms, fuzzy integrals are non linear operations based on *fuzzy measures*, which are the generalization of classical measures. In the following subsections a brief description of the theory of fuzzy measures and fuzzy integrals is provided, 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 [Klir88]. In the context of classifier combination, a fuzzy measure can express the level of competence of a classifier in assigning a pattern to a class. It must be noted that this is different from the concept of membership grade. In the latter case a value is assigned by a classifier to a pattern, expressing 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 for a given class.

Formally, a 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 CLASSIFIERS COMBINATION

By using the concepts of fuzzy measures [Klir88], a fuzzy integral has been defined [Suge77] as 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 $\mathbf{X} = \{x_1, x_2, \ldots, x_m\}$ 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 $g_k(x_i)$ is considered 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 fuzzy measures do not follow the *addition rule*, Sugeno's proposal is used to compute the fuzzy measures for the new sequence of classifiers, as shown in Eq. 1:

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

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:

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

Sugeno has proved that there is always a 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 [Taha90]:

$$e_k = \int_Y h_k \circ g_k = \max_{i=1}^{m} [h_k(x_i) \cdot g_k(y_i)] \qquad \textbf{(3)}$$

This expression is computed in two steps:

1. Obtain the product (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 for fuzzy integrals; here it is useful to see them as a method for 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.

The complete algorithm for classifiers fusion, adapted from [Taha90], is shown below in an informal way, presenting a clear view about a real world application of the previously seen concepts.

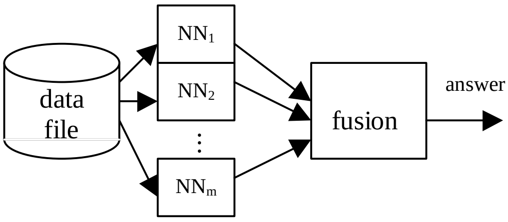
```
BEGIN classifusion,
  FOR each class tk
        FOR each classifier xi
              determine gk(xi)
        END_FOR
        compute λk
  END_FOR
  FOR each object A
        FOR each class tk
              FOR each classifier xi
                    read hk(xi)
              END_FOR
              compute the integral ek
        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 employs a set of *m* neural networks in order to label a sample Landsat image and then proceeds to an information fusion stage which provides a definitive classification answer.



Stages of the information fusion system.
Figure 1

The classification module comprises a plurarity of neural networks, which produce the inputs for the fusion stage. The values of h_k for each neural network were taken, respectively, as the direct output of the network. The heart of the fusion stage is a fuzzy integral algorithm (cf. previous sections), which works upon each individual classifier output.

3.2. THE NEURAL NETWORKS

In all classification experiments performed in this work, every neural networks has a *feed-forward* architecture with a single hidden layer.

As mentioned before, Back-Propagation is used as the learning algorithm, with adaptive learning rate and fixed momentum [Hayk98].

The patterns available for the design have been separated into three sets: the training set, the validation set, and the test set.

The networks are trained through successive epochs by using the training set as inputs. After each epoch the mean squared error (MSE) over the validation set is computed. The training goes forth for one more epoch until the MSE starts increasing.

After training the network performance is estimated by applying the testing set on the network input and computing the classification error.

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. Each pixel is represented by 24 bits, 8 bits for each channel; therefore, the network has 24 inputs. The output layer is composed of nine processors, one for each class of images.

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

3.3. LOCAL EXPERT NEURAL NETWORKS

The usual approach for neural network-based pattern recognition applications consists of training a single network so that it achieves the lowest classification error over all classes. Our approach, on the contrary, searches for the most competent network for each class, resulting in a set of *local expert networks.*

For this purpose a local competence measure is defined, as in [Ueda00]:

$$g_k(\mathrm{x}_i) = \frac{C_{kk}}{C_{kk} + \sum_{j,j\neq k} C_{kj} + \sum_{j,j\neq k} C_{jk}} \quad (4)$$

for $k = 1, \ldots, n$. In Eq. 4 C_{kk} is the number of patterns in the validation set assigned by the classifier x_i to the class t_k. Therefore $\Sigma_{j,j\neq k} C_{kj}$ is the number of validation patterns belonging to the class t_k and assigned by x_i to a different class. Similarly $\Sigma_{j,j\neq k} C_{jk}$ is the number of validation patterns not belonging to the class t_k and assigned by x_i to t_k. The so defined local competence measure will be used later in the fusion step as the fuzzy measure associated to a classifier for each class.

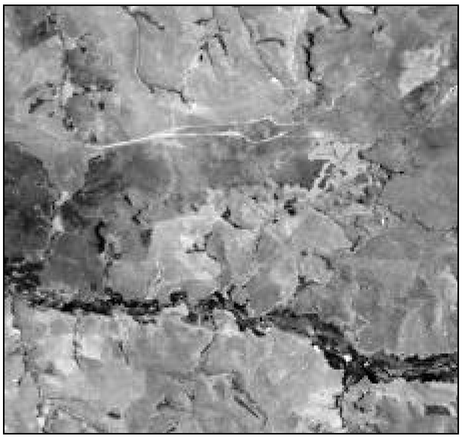
A local expert network for a class t_k can be built by introducing a small modification in the way the conventional back-propagation training algorithm computes the average output error each time a pattern is presented to the network, according to:

$$MSE_k = \frac{(a_1 - d_1)^2 + \cdots + g(a_k - d_k)^2 + \cdots + (a_n - d_n)^2}{n + g - 1} \quad (5)$$

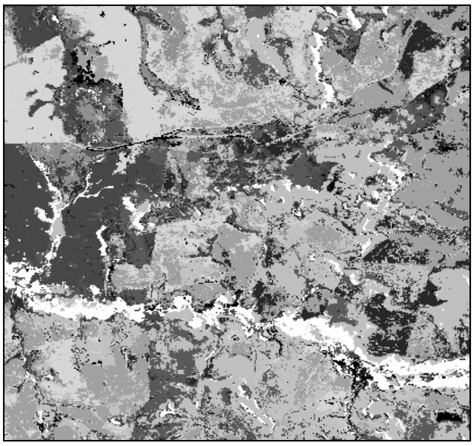
where a_i and d_i are respectively the actual and the desired values at the output *i;* and $g \geq 1$ is the expertise factor. This modification emphasizes the difference between actual and desired output for class t_k. As a consequence, the trained network will be more accurate at output *k*, corresponding to the class t_k, than at all other outputs. The improved accuracy at output *k* is obtained at the expense of the accuracy at the other outputs. The greather g the greather the accuracy at output *k* and the lower the accuracy at the other outputs. Note that for $g = 1$, equation (5) corresponds to the standard MSE.

4. EXPERIMENTS AND RESULTS

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).



"Agua-Limpa" micro-bay.
Figure 2



Thematic map used as reference.
Figure 3

The reference classification – necessary for supervised training – has been produced by an expert, mainly by using his own knowledge along with some specific GIS tool (Fig. 3).

The training and evaluation sets had respectively 300 and 30 patterns per class. Three values for γ were tested: 1 (which corresponds to a non-expert network), 10 and 15. The training went through a maximal of 800 epochs, or untill the mean squared error measured in the evaluation set started growing, whichever occured earlier.

The procedure described in section 3.3 was expected to produce a set of networks such that

$$g_k(\mathrm{x}_k) \geq g_k(\mathrm{x}_j), \quad \text{for all } j=1, \ldots, n, j \neq k \qquad (6)$$

where x_k is the local expert network for class t_k. In other words, the expert network for class t_k was expected to have the best performance for this class than any other trained network .

The experimental results obtained with the validation set confirmed this expectation for most classes: the local expert was among the 2-top and 3-top performers respectively in 67% and 78% of the cases for γ=15.

In fact there is no theoretical warranty that by training a network with emphasized error at the output corresponding to one class will always lead to the most competent network for this class. In this respect, it should be noted that the competence measure of Eq. 5 is computed over the validation set, that is, over patterns not used in the training phase. Nevertheless the experimental results have shown that the error-emphasis procedure does "tend" to improve the performance of a network for a particular class.

The procedure described in section 3.3 will produce a set of networks with different competences for each class. Thus one can select among them the most competent networks for each class to be used in the fusion phase. This was the procedure followed in this work. The best performance networks for each class were chosen among all networks produced by using the error-empahsis approach, with γ=1, 10 and 15. Table 1 shows the competence measures of these selected networks. The numbers in the first column identify a network and correspond to the class for which it has the best performance, in accordance with Eq. 5. Notice that the diagonal of the data in Table 1 contains the best values along each column. It is worth mentioning that this procedure does not necessarily lead to *n* (the number of classes) different networks. Actually rows 1, 6 and 7 in Table 1 correspond to the same network.

For comparison purposes the competence measures of the conventional non-expert network (γ=1) is presented in Table 2.

Table 3 shows the recognition rate per class for the combined classifier and for the conventional non-expert network computed over the entire image. If one considers all patterns in the image, this rate is similar to the competence measure defined in Eq. 5, whereby the number of patterns falsely assigned to t_k in the denominator is discarded.

It is interesting to note the important performance improvement obtained by the combination for the classes with the poorest performance in the conventional network, namelly classes 8 and 9.

The last column at right shows the average value along each row. It represents the average recognition rate for an image containing the same number of pixels for each class. The combination of neural networks through fuzzy integrals was able to increase the average recognition rate from 87% to 91%.

Neural Network	Competence Measure per Class									
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Average
1	**0.94**	0.77	0.76	0.72	0.81	0.94	0.97	0.90	0.47	0.81
2	0.88	**0.82**	0.79	0.76	0.81	0.82	0.93	0.84	0.61	0.81
3	0.71	0.50	**0.88**	0.44	0.73	0.70	0.74	0.77	0.46	0.66
4	0.88	0.67	0.69	**0.83**	0.85	0.78	0.84	0.85	0.59	0.78
5	0.82	0.74	0.82	0.68	**0.88**	0.74	0.83	0.85	0.59	0.77
6	0.94	0.77	0.76	0.72	0.81	**0.94**	0.97	0.90	0.47	0.81
7	0.94	0.77	0.76	0.72	0.81	0.94	**0.97**	0.90	0.47	0.81
8	0.88	0.61	0.71	0.56	0.76	0.76	0.94	**0.94**	0.56	0.75
9	0.88	0.66	0.61	0.00	0.74	0.68	0.68	0.90	**0.63**	0.64

Competence Measure of the Networks Selected for Fusion
Table 1

Network	Competence Measure per Class									
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Average
Conventional	0.94	0.77	0.76	0.72	0.81	0.94	0.97	0.90	0.47	0.81

Competence Measure of the Conventional Non-Expert Network
Table 2

Classifier	Recognition Rate									
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Average
Conventional	0.98	0.93	0.88	0.94	0.88	0.92	0.92	0.77	0.53	0.87
Combined	0.98	0.88	0.91	0.93	0.87	0.92	0.94	0.91	0.83	0.91

Recognition Rates for the Conventional Network and for the Combined Networks
Table 3

5. CONCLUSION

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

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

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

Results encourage a deeper research of combined classifiers for this kind of application as well as the procedure to produce expert net networks. Among the open questions to be further investigated is how the value of expertise factor γ innfluences the competence measure of the resulting network.

REFERENCES

[Hayk98] – S. Haykin - *Neural Networks – A Comprehensive Foundation*, Mcmillan College Publishing Co (1998).

[Kitt98] - 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).

[Klir88] – G. Klir, T. Folger, - *Fuzzy Sets, Uncertainty and Information*, Prentice-Hall International, Inc (1988).

[Lill00] - T. M. Lilesand, R. W. Kiefer, *Remote Sensign and Image Interpretation*, 4th Ed., Wiley, (2000).

[Math99] - P. M. Mather, - *Computer Processing of Remotely-Sensed Images: An Introduction*, 2nd Ed, Willey (1999).

[Rich99] – J. A. Richards, X. Jia - *Remote Sensing Digital Image Analysis – An Introduction*, 3rd Ed., Springer Verlag (1999).

[Suge77] M. Sugeno - *Fuzzy measures and fuzzy integrals: A survey*, Fuzzy Automata and Decision Processes. Amsterdam: North Holland, pp.89-102 (1977).

[Taha90] - 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).

[Ueda00] - Ueda, N., *Optimal Linear Combination f Neural Networks for Improving Classification Performance*, IEEE Transactions Pattern Analysis and Machine Intelligence, Vol. 22, NO. 2; 207-214 (February 2000).