A Comparison of Gaussian Based ANNs for the Classification of Multidimensional Hyperspectral Signals

A. Prieto, F. Bellas, R. J. Duro, F. Lopez-Peña

Grupo de Sistemas Autónomos, Universidade da Coruña, Spain
abprieto@udc.es, fran@udc.es, richard@udc.es, flop@udc.es

Abstract. This paper is concerned with the comparison of three types of Gaussian based Artificial Neural Networks in the very high dimensionality classification problems found in hyperspectral signal processing. In particular, they have been compared for the spectral unmixing problem given the fact that the requirements for this type of classification are very different from other realms in two aspects: there are usually very few training samples leading to networks that are very easily overtrained, and these samples are not usually representative in terms of sampling the whole input-output space. The networks selected for comparison go from the classical Radial Basis Function (RBF) network to the more complex Gaussian Synapse Based Network (GSBN) considering an intermediate type, the Radial Basis Function with Multiple Deviation (RBFMD). The comparisons were carried out when processing a benchmark set of synthetic hyperspectral images containing mixtures of spectra from materials found in the US Geological Service database.

1 Introduction

As remote sensing of the earth becomes a more important element in the toolbox of scientists, the mechanisms for processing the ever larger data streams coming from the instruments used for this purpose increase their relevance. High resolution hyperspectrometers are among the instruments that will be used in an ever increasing number [1][2][3]. In hyperspectral remote sensing images, each spatial resolution element data is acquired with high spectral resolution over the electromagnetic spectrum ranging the 400-2500 nm (visible to near infrared). It is commonplace to use 50 to 250 spectral bands of bandwidths in the 5 to 20 nm range. The large amount of information hyperspectral imaging provides permits a detailed description of the spectral signature thus greatly improving the ability to detect and identify individual materials or classes with respect to other remote sensing techniques.

Any analysis or classification method for hyperspectral image processing is aimed at the identification of what pixels contain different spectrally distinct materials (endmembers) and in what proportions. A number of approaches based on statistical theory or using filtering or correlations have been applied to the analysis of these data sets by different authors with the objective of improving the classification results [4]. Some of these methods are compared using classification performance in [5].

From this point of view, artificial neural networks (ANNs) appear as a very suitable and effective alternative to deal with spectral image analysis difficulties [6] [7]. In the

case of ANNs trained using supervised training algorithms, the main drawback to performing the analysis and classification of hyperspectral remote sensing data is the difficulty in obtaining labelled sample data that can be trusted to the point of using it for quantitative evaluation. The scarcity of ground truth data has been recognized and specific training strategies have been devised to cope with this handicap [8], [9]. In fact, in an ideal world, we would be able to obtain a training procedure that produces good classifiers from very small training sets. This problem becomes even more pronounced when pixels correspond to combinations of materials, that is, the "spectral unmixing" problem [6]. Ideally, hyperspectral images may allow the identification of single pixel materials. However as these pixels are frequently combinations of materials, it is often a plus to be able to decompose each pixel spectrum into its constituent material spectra. The cause of the mixture of several material spectra into a hyperspectral pixel spectrum may be that different land covers are included in the area whose radiance measurement results in an image pixel.

In this work we assume a linear mixture model, in which several basic materials (endmembers) are combined according to some abundance coefficients at each image pixel. Taking its spatial distribution, the abundance coefficients may be visualized as abundance images, which provide a description of the spatial distribution of the material. The computation of the abundance coefficients given a pixel spectrum and a set of endmembers is what is termed the unmixing procedure. If the endmembers are given, the unmixing procedure is equivalent to the parallel detection of the spectral features represented by the endmembers.

One basic approach is classical, if you concentrate only on what is relevant the classification becomes much more robust and efficient. This is the approach followed in the work leading to this paper. An Artificial Neural Network architecture and training algorithm that implement an automatic procedure to concentrate on what is relevant and ignore what is not straight from the training set is required in order to effectively perform the task. To do this, many authors have resorted to Gaussian based ANNs as a way to implicitly establish this filtering ability in the network. This is the case of the work of Dundar and Landgrebe [10] with RBFs, where the authors claim that Gaussian RBFs have proven to be in their case the most effective network for hyperspectral image processing, or that of Crespo et al. [11] with GSBNs where the Gaussian processing has been transferred to the synapses, thus providing more degrees of freedom.

In this paper the objective is to evaluate the performance of three Gaussian based types of ANNs when dealing with multidimensional signals and very few training points are available. These three types of networks range from the classical Radial Basis Function Network as proposed above, to a modification of RBFs where they have been endowed with trainable deviations for each dimension, to the more versatile Gaussian Synapse Based networks. In the sections that follow we will describe this networks and through a similar backpropagation based training algorithm we will compare their performance when unmixing a benchmark set of images based on Graña et al´s repository.

2 Description of the ANNs

Gaussian based ANNs have been widely applied in image processing due to their capabilities in noise filtering. As commented in the previous section, we are going to compare the results provided by three different types of these networks. In the radial basis function (RBF) neural networks the input layer directly transmits the inputs to the neurons of the hidden layer. In these neurons the Gaussian function is applied over the inputs by using one parameter per synapse (center) and one parameter per hidden neuron (deviation). The output provided by the neurons in the hidden layer passes to the output neurons trough a linear combination of weights as in a perceptron. The radial basis function with multiple deviation (RBFMD) neural networks are structurally similar to RBF networks but having one deviation parameter per synapse instead of per neuron in the hidden layer. Finally, the Gaussian synapse based networks (GSBN) have a multilayer perceptron structure, but replacing the simple weights used in the synapses by Gaussian functions in order to filter the inputs. The main difference between these three types of networks is the number of parameters to be trained, being the RBF the one requiring fewer parameters and the GSBN the one requiring most. As it is shown in

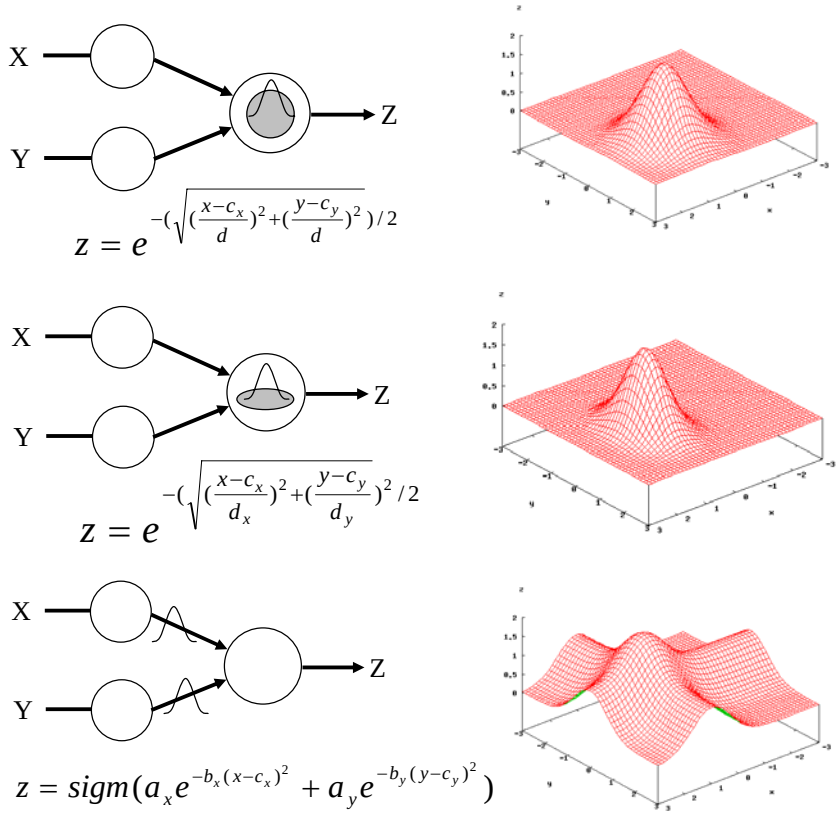

Fig. 1. Top graph represents an RBF neural network, middle graph and RBFMD neural network and bottom graph a GSBN.

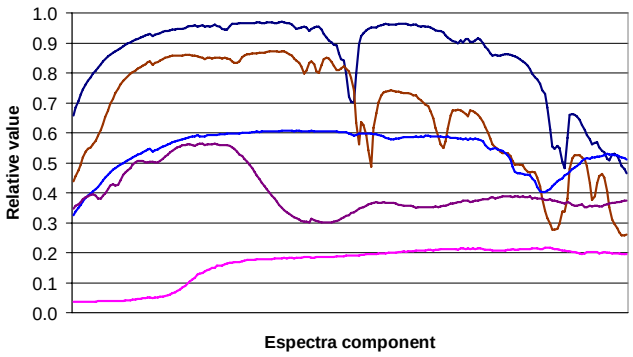

Fig. 2. Representation of the spectra corresponding to the 5 endmembers used to generate the training and test data sets.

the next section, this number of parameters determines the complexity of the learning process and the generalization capability.

The relevant differences between these types of ANNs arise from their basic structural units as shown in Fig. 1. These units have two input neurons (x and y) and one output neuron (z). The top graph of Fig. 1 corresponds to a RBF neural network, in which a single parameter per synapse (the centers C_x and C_y) should be trained. The middle graph corresponds to a RBFMD neural network where two parameters per synapse must be trained (the centers C_x and C_y and the deviations d_x and d_y). Finally, at the bottom is a GSBN neural network where 3 parameters per synapse should be trained (a_x, b_x, c_x, a_y, b_y and c_y). The right part of Fig. 1 provides a representation of the decision boundaries implementable by each of these three ANNs when the input variables range between -3 and 3. This corresponds to the possible outputs provided by a certain neuron

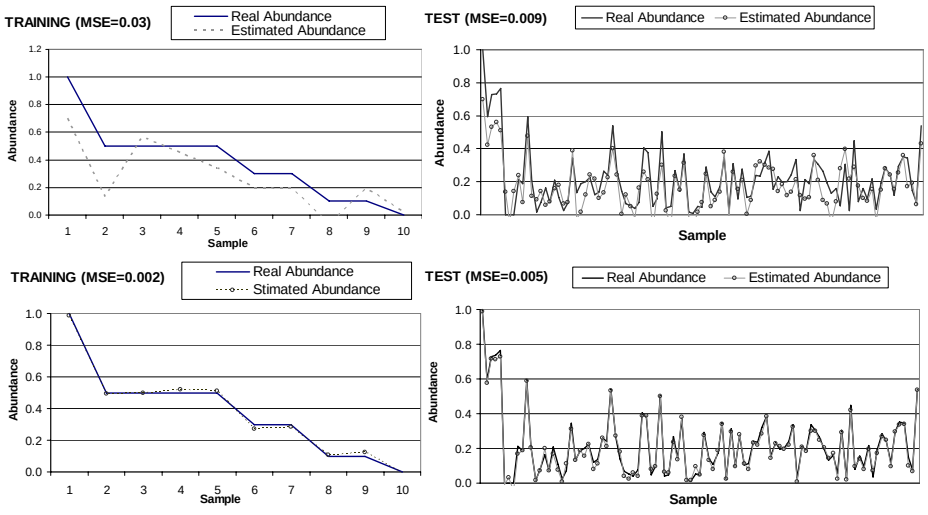

Fig. 3. Modeling over the training set (left) and the test set (right) provided by one of the networks for two significant error values: 0.009 in top graphs and 0.005 in bottom graph.

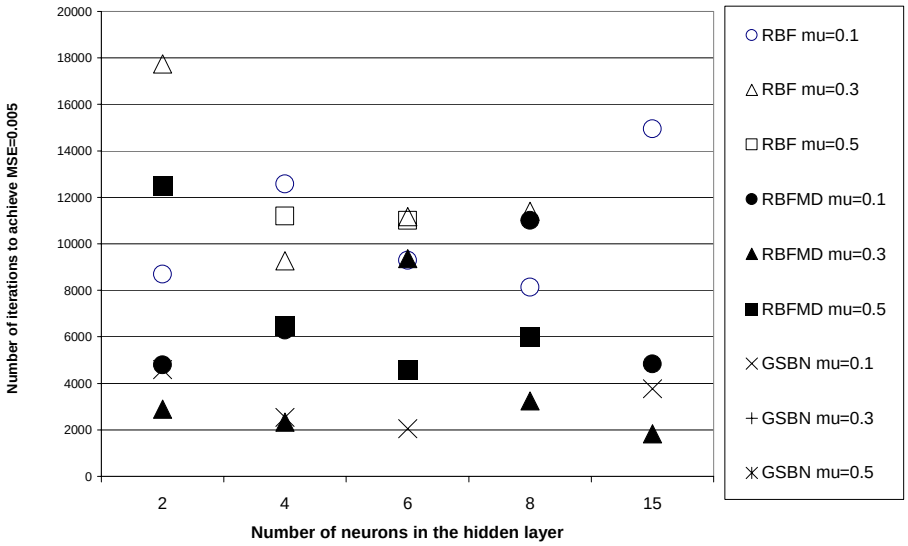

Fig. 4. Number of neurons in the hidden layer versus the number of iterations to achieve the desired for the 3 types of networks (RBF represented by hollows, RBFMD by solids and GSBN by marks) with 3 different mu values for each one

of the hidden layer to the next layer. As it can be seen, RBF networks (top graph) provide symmetric decision boundaries in both axes while the RBFMD ones (middle graph) are able to filter independently on each axis presenting a more flexible decision boundary. This capability is very suitable for our purposes because the network has higher degree of freedom to filter the information. In the case of the GSBN network (bottom graph) we can see a more complex decision boundary because the filtering appears in both axes independently again and, in addition, the shared zone could be discriminated through the value of the sigmoids in the neurons.

The main reason for studying networks that provide more complex and flexible decision boundaries is the high dimensionality of the search space and the complexity of the subspace to discriminate in hyperespectral unmixing problems. The application of these networks increases the number of parameters that must be trained and, consequently, the computational cost but it should be compensated by the decrease of the minimum necessary network size and, what is more important, the speed of the network training stage as well as the need to use fewer training samples.

In the next section we will show the result of the comparison between the three types of networks presented applied to a hyperespectral image.

3 Comparison of networks

To compare the networks we have used a set of 5 synthetic endmembers developed by Graña et al. [14] which is shown in Fig. 2. A training set containing 10 spectra resulting from the linear combination of these 5 endmembers and a test set of 1000 spectra have

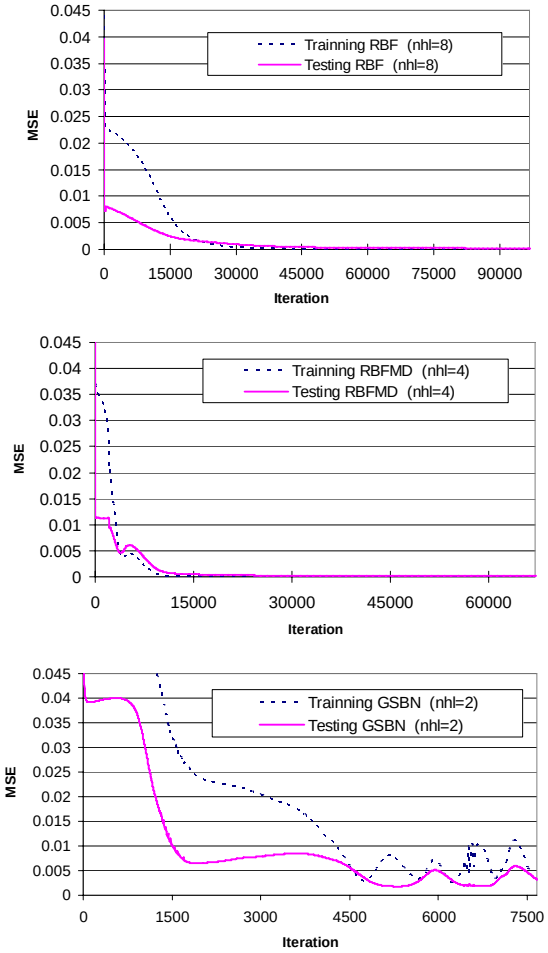

Fig. 5. MSE evolution with respect to the number of iterations in the training process and in the test process using a RBF network with 8 neurons (top graph), a RBFMD with 4 neurons (middle graph) and a GSBN with 2 neurons (bottom graph).

been created. The training algorithms applied are variations of the classical backpropagation algorithm modified for the radial basis functions [13] and for the Gaussian synapses [12]. This training process is applied over the training set for each type of network and every iteration the network is tested over the test set.

First of all, in Fig. 3 we have represented the modeling over the training and the test sets provided by one of the networks for two significant error values. From the curves we can establish that the learning has been successful when the MSE is lower than 0.005 according to the modeling over the test set.

To measure the quality of the network obtained we will focus on two main parameters: the number of iterations required to achieve a given error value and the stability of error evolution (to detect overtraining problems). In the first case, in order to compare the three types of networks we have used two parameters: the learning

coefficient *mu* and the number of neurons in the hidden layer. In Fig. 4 we show number of neurons in the hidden layer versus the number of iterations to achieve the desired error for the 3 types of networks (RBF represented by hollow symbols, RBFMD by solid symbols and GSBN by line symbols) and with 3 different *mu* values for each one. The data shown in the figure provide a good indicator of several facts. On one hand, RBF based networks are a lot slower than other types of networks in every case, even when taking into account the same number of parameters. Take into account, that an RBF with 6 neurons has the same number of trainable parameters than a RBFMD with 3 neurons and a GSBN with 2 neurons. The RBFMD is clearly more efficient in terms of reaching the solution in less iterations, but the one that performs the best is the GSBN. This is so for just about any size of the network. The figure shows that the influence of the *mu* parameter is only significant in the case of GSBN networks where just low *mu* values provide valid errors, if training is carried out with high *mu* the networks take very long to achieve a result and often they will end up overtraining..

In terms of overtraining, taking into account the different number of parameters that characterize each type of network, we must select networks with different number of neurons in the hidden layer in order to perform an equivalent comparison. Thus, in Fig. 5 we have represented the MSE evolution against the number of iterations in the training process and in the test process using a RBF network with 8 neurons (top graph), a RBFMD with 4 neurons (middle graph) and a GSBN with 2 neurons (bottom graph). In this test, we have provided a little advantage for the RBF in terms of number of parameters, but we wanted to make sure it did not overtrain. As we can see, in the case of the RBF network, the evolution of the error in the test process is highly stable because the network does not overtrain. On the other hand, GSBN and, mainly, the RBFMD networks are more unstable they are very sensitive to overtraining and having achieved a given error for the test set as training progresses this level of error may increase and even create large oscillations in the case of RBFMD with large *mus*..

This way, from these results the main conclusion we can extract is that although the three types of networks achieve the desired error value (MSE=0.005) in almost all cases, the RBF networks are the one that need more iterations for training for the same number of trainable parameters. The GSBN networks are the ones that need less iterations. Finally, RBFMD are in an intermediate point in both features but are quite sensitive to overtraining, probably because of the deviation parameter in the denominator of the exponential.

Conclusions

In this paper we have compared three types of Gaussian based artificial neural networks in terms of their performance when addressing the spectral unmixing problem in a set of benchmark synthetic hyperspectral images. From this comparison it is clear that for equivalent number of parameters in the networks, the Gaussian synapse based networks converge to very good classification results in less iterations than the other two, whereas the radial basis functions do so in a much smoother fashion. In terms of the overtraining behavior, it is clear to see that the RBF networks are much harder to overtrain and thus perform much better on the test sets independently of the length of the training period

(which in many cases is very hard to set beforehand), obviously, at the cost of a much slower training process.

Acknowledgements

This work was funded by the MCYT of Spain through project VEM2003-20088-C04-01 and Xunta de Galicia through project PGIDIT03DPI099E.

References

1. Fingas, M.F. and Brown, C.E.: Review of Oil Spill Remote Sensing, in Proceedings of the Fifth International Conference on Remote Sensing for Marine and Coastal Environments, Environmental Research Institute of Michigan, Ann Arbor, Michigan, pp. I211-218. 2000
2. Regional Marine Pollution Emergency Response Centre for the Mediterranean Sea (REMPEC): MEDIPOL 2003 - European workshop on satellite imagery and illicit oil spills in Europe and in the Mediterranean. Final Recommendations, 2003
3. Bissett, W.P., Arnone, R., Davis, C.O., Dickey, T., Dye, D.R., Kohler, D.D.R., Gould, R.: From meters to kilometers - a look at ocean color scales of variability, spatial coherence, and the need for fine scale remote sensing in coastal ocean optics. Oceanography, 2003
4. Campbell, J.: Introduction to Remote Sensing. The Guilford Press ,1996
5. Hsu, P.H. and Tseng, C.E.: Feature Extraction for Hyperspectral Image. Proceedings of the 20th Asian Conference on Remote Sensing, Vol. 1, 405-410, 1999
6. Merényi, E., Minor, T. B., Taranik, J. V., and Farrand, W.H.: Quantitative Comparison of Neural Network and Conventional Classifiers for Hyperspectral Imagery, In Summaries of the Sixth Annual JPL Airborne Earth Science Workshop, Pasadena, CA, March 4-8, 1996, Vol. 1: AVIRIS Workshop, Ed. R.O.GreeN, 1996.
7. Ghosh, J.: Adaptive and neural methods for image segmentation. In Al Bovik, editor, Handbook of Image and Video Processing, chapter 4.10, pages 401--414. Academic Press, 2000.
8. Tadjudin, S. and Landgrebe, D.: Covariance Estimation with Limited Training Samples, IEEE Trans. Geos. Rem. Sensing, 37(4) , 2113- 2118, 1999
9. Tadjudin, S. and LandgrebeRobust, D.: Parameter estimation for mixture model, IEEE Trans. Geos. Rem. Sensing, 38(1): ,439, 2000
10. M. Murat Dundar and David Landgrebe: Toward an Optimal Supervised Classifier for the Analysis of Hyperspectral Data IEEE Transactions on Geoscience and Remote Sensing, Volume 42, No. 1, pp 271-277, January, 2004.
11. J. L. Crespo, R.J. Duro, F. López Peña: Unmixing Low Ratio Endmembers through Gaussian Synapse ANNs in Hyperspectral Images Proceedings CIMSA 2004 IEEE International Symposium on Computational Intelligence for Measurement Systems and Applications. V1, pp: 150 -154, 2003
12. Duro, R.J., Crespo, J.L., and Santos, J.: Training Higher Order Gaussian Synapses. LNCS, Vol. 1606 Springer-Verlag, Berlín, 537-545, 1999
13. Karayiannis, N.B.: "Reformulated radial basis neural networks trained by gradient descent" IEEE Transactions on Neural Networks, vol. 10 no. 3 , 657 –671, 1999
14. Graña, M., Raducanu, B., Sussner P., and Ritter. G.: On Endmember Detection in Hyperspectral Images with Morphological Associative Memories. Presented at IBERAMIA 2002, Sevilla, Spain, 526-535, 2002