Neural-Network Combination for Noisy Data Classification

Fabricio A. Breve, Moacir P. Ponti Jr., Nelson D. A. Mascarenhas
Departamento de Computação, UFSCar – Universidade Federal de São Carlos
Rod. Washington Luís, Km 235, Caixa Postal 676, 13565-905
São Carlos-SP13566-050, São Carlos-SP, Brazil.
Tel.: +55-16-33518579 / 3351-8232 – Fax.: +55-16-3351-8233
E-mail: {fabricio,moacir,nelson}@dc.ufscar.br

Abstract

Classifier combination experiments using neural network-based classifiers were carried out using noisy soil science multispectral images, which were obtained using a tomograph scanner. Using few units in the hidden layer images were classified by the Multilayer Perceptron (MLP) and the Radial Basis Function Network (RBF). Later we used classifier combining techniques as Decision Templates (DT) and Dempster-Shafer (DS), in order to improve the performance of the single classifiers and also to stabilize the performance of the Multilayer Perceptron. Classification results were evaluated through the Estimated Error (by the Hold-Out technique), and the Kappa Coefficient. The results showed that the RBF Network obtained a small improvement in performance with the combination. However, we observed a good improvement in stabilization of the Multilayer Perceptron, especially with the Decision Templates method.

1. Introduction

There are many techniques for combining multiple classifiers. They appeared on literature mainly in the past 20 years. The idea behind combiners is that different individual classifiers can offer complementary information about the objects to be classified. Instead of using just one classifier, a safer option would be to use many classifiers and combine their outputs [2]. The combination of classifiers has the intuitive purpose of improving performance, especially on challenging problems like handwriting recognition and others [3].

In a previous work [4], good results were obtained using a combination scheme for the Multilayer Perceptron classifier showing some increase in performance. In this paper we present a set of experiments in order to recognize materials in noisy soil science multispectral images. These images were classified by neural network based classifiers: Multilayer Perceptron and Radial Basis Function - and classifier combining techniques - Decision Templates and Dempster-Shafer - were investigated. We also present a performance comparison between the individual classifiers and the combiners. The results were evaluated by the estimated error, obtained using the Hold-Out technique, and the Kappa coefficient.

The text is organized as follows. In Section 2 we introduce the image acquisition scheme. In Section 3 we present the classifiers and combiners used in this paper. Section 4 describes how the methods were evaluated. Section 5 explains how we conducted the experiments. Finally, Section 6 shows our conclusions and final remarks.

2. Image Acquisition

The computerized tomograph (CT) scanner used to acquire the images is a first generation equipment developed by Embrapa[1] to explore applications in soil science. It has fixed X and γ-ray sources, while

[1] Brazilian Agricultural Research Corporation – state institute for scientific research on agriculture

the object being studied is rotated and translated. All the system hardware and software was developed by Embrapa. [5]

In this work we used images of a phantom containing four materials commonly found in soil: calcium, phosphorus, water and aluminum. The phantom has a cylindrical base of plexiglass (polymer), and has four cylinders inside, each one containing one of the materials, as shown in Figure 1.

The images were obtained using two X-ray sources and two Γ-ray sources (Cesium and Americium). The X-ray energies were 40keV and 85keV. The Γ-ray sources were 662keV (Cesium) and 60keV (Americium).

The images have a resolution of 65x65 pixels, and were obtained using only 3 seconds of exposure. After the reconstruction with the filtered backprojection algorithm they were normalized to 256 levels of gray, which are proportional to the values of the physically observed linear attenuation coefficients.

The four images are shown in Figure 2. Together they compose a multispectral image, which is the object of study of this work.

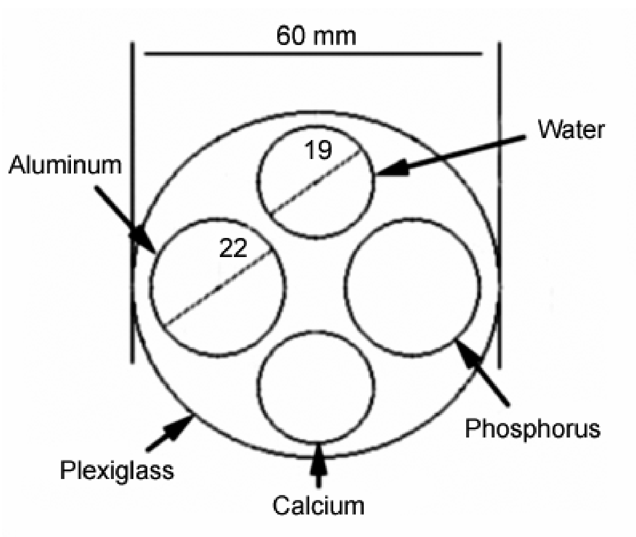


Figure 1. Phantom construction diagram with dimensions and materials

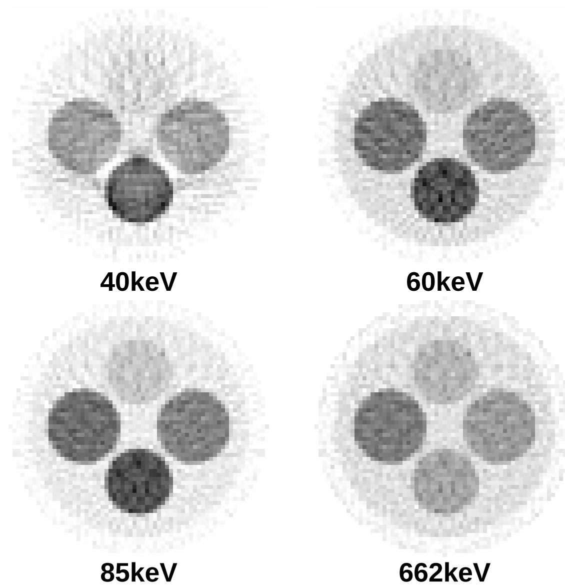


Figure 2. Multispectral image bands acquired by an X and γ-ray CT scanner with multiple energies: 40keV, 60keV, 85keV and 662keV

3. Classification Methods

Classification was first performed by using individual Multilayer Perceptron and Radial Basis Function classifiers. Later, Bagging technique, Decision Templates and Dempster-Shafer classifier combiners were used in order to improve the performance of the single classifiers.

3.1 The Multilayer Perceptron

This kind of neural network is composed by a set of sensorial units organized in three or more layers. The first layer is the *input layer*, which does not perform any computational task. There are one or more *hidden (intermediate) layer*s and an *output layer*, all composed by computational nodes.

This kind of neural network has been used with success to solve difficult problems through its training by using the error backpropagation algorithm, which basically consists of two steps: a forward step where the signal propagates through the computational units until it gets to the output layer; and a backwards step where all the synaptic weights are adjusted accordingly to an error correction rule. [6] [7]

3.2 Radial Basis Functions

This kind of network uses an approach where the neural networks are seen as a problem of curve adjustment. In this context, learning is finding a surface in a multidimensional space that provides the best adjustment for the data in a statistical point of view.

Radial Basis Networks have three layers with totally different roles. The first layer is composed by sensorial nodes that receive data from the environment. The second layer, the only hidden layer in the network, performs a non-linear transformation from the entry-space to a high-dimensional hidden-space. Finally, the output layer is linear and provides the network answer to an input signal. [6][7]

3.4 Decision Templates

When using classifiers that give us continuous-valued outputs (like MLP and RBF) we can treat the outputs as confidences in proposed labels and estimates of the posterior probabilities for each class.

In this combiner the idea is to organize each of these outputs in a matrix called Decision Profiles (DP). We can take the most common DP for each class and call it the *Decision Template* (DT) of that class. Then, when we want to classify a given sample we build its DP and compare it with the DT of each class using some measure distance (like the Euclidean distance). The closest match will label the sample. [2]

3.5 Dempster-Shafer

Dempster-Shafer (DS) is based on the Evidence Theory, proposed by Glen Shafer as a way to represent cognitive knowledge. In this formalism, the best probability representation is a belief function, rather than a Bayesian distribution. Probability values are assigned to a set of possibilities instead of unique events. Its appeal is in the fact that they code evidences rather than propositions. It provides a simple method of combining evidences from different sources (Dempster rule) without any a priori distribution. [8]

The training algorithm for DS is the same algorithm used to train the DT combiner, where the DT´s for each class are found from the training data. The difference here is that instead of calculating the similarity between the DP of a given sample and each DT, we calculate the closeness between the DT and the output of each classifier. These closeness values are used to calculate a belief degree for each classifier for each one of the classes. The final degrees of support for each class are calculated from the belief degrees.[2]

4. Evaluation

The performance of the classifiers and the combiners were evaluated by using the Estimated Error (by Hold-Out technique) and the Kappa Coefficient.

4.1 Hold-Out

In this method we split the available data in two sets, one for training the classifier and the other one for testing it, obtaining the error rate. [2]

This is a pessimistic method because it only uses half of the available data for training [9]. However, there is a good reason to use it to evaluate neural networks: Hold-Out is a fast evaluation method and neural networks classifiers are much slower to train than statistical classifiers.

4.2 Kappa Coefficient

The Kappa coefficient (K) can be used to measure the agreement rating between two classifiers. To evaluate the performance of a classifier we can use Kappa to compare the output of a classifier with the pre-labeled samples [10].

Kappa gives us a value in the [-1 1] interval. When the agreement is no higher than what is expected in a random classification, K will be 0 and when there is full agreement between the sample labels and the classifier output K will be 1. So, the higher is the value of K, the best is the expected performance of the classifier. The interpretation of the Kappa values is subjective and depends on the level of correctness required by the problem. [11] [12]

5. Experiments

We took 80 samples in 10x8 pixels windows from each of the 6 classes (water, aluminum, phosphorus, calcium, plexiglass and background) in a total of 480 samples. We split this set in two halves, following the hold-out technique, so there would be 240 samples (40 for each class) for training and another 240 samples for testing. We used the four available bands, so we have 4 features, which also means that the input layer of our networks have 4 nodes.

The MLP networks we trained had from 2 to 10 units in one single hidden layer, and the RBF networks had from 2 to 10 units in the hidden layer as well. These numbers were selected because there is no foolproof way to tell a priori how many units in the hidden layer would be the best choice. [13]

Table 1. Estimated Error for the Multilayer Perceptron

Units	Single	DT	DS
2	0.5720	0.0349	0.0613
3	0.2689	0.0163	0.0275
4	0.1318	0.0141	0.0177
5	0.0976	**0.0123**	0.0151
6	0.0741	0.0127	0.0139
7	0.0681	0.0129	0.0138
8	0.0636	0.0130	**0.0137**
9	**0.0511**	0.0134	0.0138
10	0.0570	0.0135	0.0139

Table 2. Kappa Coefficient for the Multilayer Perceptron

Units	Single	DT	DS
2	0,3137	0,9581	0,9265
3	0,6773	0,9805	0,9671
4	0,8419	0,9831	0,9788
5	0,8829	**0,9853**	0,9819
6	0,9111	0,9848	0,9833
7	0,9183	0,9845	0,9835
8	0,9237	0,9844	**0,9836**
9	**0,9387**	0,9840	0,9835
10	0,9316	0,9838	0,9833

Table 3. Standard Deviation for the Multilayer Perceptron

Units	Single	DT	DS
2	0,2554	0,0286	0,0348
3	0,2892	0,0061	0,0133
4	0,2073	0,0044	0,0065
5	0,1504	0,0029	0,0046
6	0,1199	0,0020	0,0028
7	0,1065	**0,0013**	0,0024
8	0,0842	0,0014	**0,0020**
9	**0,0777**	0,0018	0,0021
10	0,0807	0,0018	0,0023

Table 4. Estimated Error for the Radial Basis Function

Units	Single	DT	DS
2	0,5125	0,4458	0,4333
3	0,3958	0,3583	0,3583
4	0,2958	0,2583	0,2583
5	0,2292	0,2333	0,2333
6	0,1333	0,1292	0,1292
7	**0,0750**	**0,0708**	**0,0708**
8	0,0833	0,0833	0,0833
9	**0,0750**	0,0750	0,0750
10	0,0833	0,0875	0,0875

Table 5. Kappa Coefficient for the Radial Basis Function

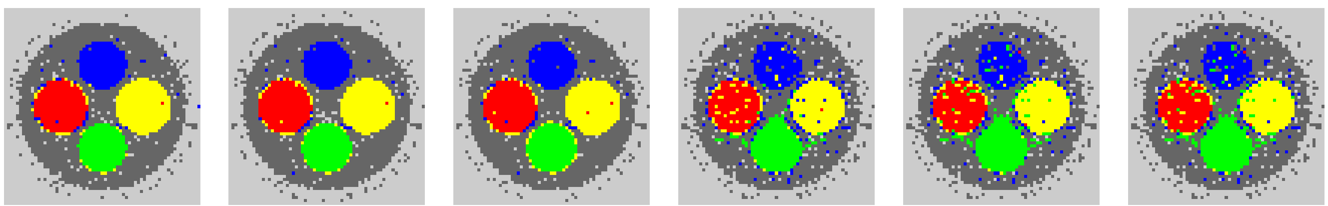
Units	Single	DT	DS
2	0,3850	0,4650	0,4800
3	0,5250	0,5700	0,5700
4	0,6450	0,6900	0,6900
5	0,7250	0,7200	0,7200
6	0,8400	0,8450	0,8450
7	**0,9100**	**0,9150**	**0,9150**
8	0,9000	0,9000	0,9000
9	**0,9100**	0,9100	0,9100
10	0,9000	0,8950	0,8950

Table 6. Standard Deviation for the Radial Basis Function

Units	Single	DT	DS
2	0,0000	0,0000	0,0000
3	0,0000	0,0000	0,0000
4	0,0000	0,0000	0,0000
5	0,0000	0,0000	0,0000
6	0,0000	0,0000	0,0000
7	0,0000	0,0000	0,0000
8	0,0000	0,0000	0,0000
9	0,0000	0,0000	0,0000
10	0,0000	0,0000	0,0000

MLP based classifiers are highly unstable and tend to present different results depending on its initialization parameters, so every experiment was repeated 100 times (including the experiments with combiners) and the values showed in this article are the mean values obtained from all these rounds. Although RBF is not unstable like MLP, we also repeated the experiments that used RBF 100 times to follow an outline. The Nguyen-Widrow [14] initialization algorithm was used to initialize the MLP networks. Adaptive learning rates were used with 0.01 as the initial value for learning rate, 1.05 as the multiplier for increasing learning rate, 0.7 as the multiplier for decreasing learning rate, 0.95 as the momentum constant and 1.04 as the error ratio.

The MLP results can be viewed in Tables 1 and 2. Since MLP is an unstable classifier we also included standard deviation results in Table 3. The RBF results can be viewed in Table 4 and 5. We also include the standard deviation in Table 6. The best results in each column are in boldface. In Figure 3 we show some thematic images for each classifier and each combination method, in order to illustrate the classification.



MLP - Single   MLP - DT   MLP - DS   RBF - Single   RBF - DT   RBF - DS

Figure 3. Some thematic images for each classifier and each combination method

6. Conclusions

In the experiments with the MLP single classifier we noticed that the results got better as we added units to the hidden layer. Using combination, the best results were achieved using fewer units in the hidden layer. The use of classifier combiners led to better and more stable classifiers, as we can see in Tables 1, 2 and 3. Dempster-Shafer and Decision Templates combiners showed good results no matter how many units there were in the hidden layer, so they should be good choices of combiners when it is not viable to conduct experiments to find out which is the optimal number of units in the hidden layer for a particular problem. They also showed that the differences between classifiers produced by different MLP initializations are enough to produce good combinations, and then bagging techniques becomes unnecessary. Both methods showed improvements over the single classifier, but Decision Templates outperformed Dempster-Shafer with all the configurations, so we would highly recommend it for MLP-based classification systems.

The experiments with RBF showed that both combining methods led to only slightly better classification, as we can see in Tables 4 and 5. That is probably due to the more stable behavior of RBF, which leads to similar classifiers when we change only the initialization. So here the classifiers do not differ from each other and it is difficult to obtain a good combination. In this case, it is hard to point out which one had the best performance. The DS method performed better than DT just in the experiments using 2 units in the hidden layer.

New advances in this field could be reached with the use of mechanisms to eliminate redundant classifiers, constrained weak classifiers and adaptive combination. A contextual approach for the classifier combination methods is also another good research way.

The results shows that using Neural Network based classifiers to identify materials on CT images is viable, even when applied to images with high noise levels. The use of classifiers combiners led to better classification and more stable MLP systems, minimizing the effects of the unstable nature of the individual MLP classifiers.

7. Acknowledgements

We would like to thank Dr. Paulo E. Cruvinel for providing the multispectral images used in the experiments, CAPES and FAPESP (grant n. 04/05316-7) for student scholarship. This work was also partially supported by FAPESP Thematic Project 2002/07153-2.

8. References

[1] Roux, C.; Udupa, J.K.; Scanning the Issue - Special Issue on Emerging Medical Imaging Technology. Proceedings of the IEEE. Volume 91, Issue 10, Oct. 2003 Page(s):1479 – 1482

[2] Kuncheva, L.. "Combining Pattern Classifiers", *Wiley-Interscience*, Hoboken, NJ, 2004.

[3] Xu, L., Krzyzak, A., Suen, C. Y. 1992. Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition, IEEE Transactions on Systems, Man and Cybernetics 22 3: 418-435.

[4] Breve, F. A.; Ponti Jr, M. P.; Mascarenhas, N. D. A., Combining Methods to Stabilize and Increase Performance of Neural Network-Based Classifiers . In: Proceedings of SIBGRAPI 2005, XVIII Brazilian Symposium on Computer Graphics and Image Processing. Natal, Brazil: IEEE CS Press, 2005.

[5] Cruvinel, P. E., Cesareo, R., and Mascarenhas, S., X and γ- Rays Computerized Minitomograph Scanner for Soil Science, IEEE Transactions on Instrumentation and Measurements, v. 39, n. 5, 1990. pp. 745-750.

[6] Haykin, S. "Redes Neurais – Princípios e Prática", *Bookman*, 2 ed. Porto Alegre, 2001.

[7] Bishop, C. M., "Neural Networks for Pattern Recognition", *Oxford,* New York, 1995.

[8] Ahmadzadeh, M. R., Petrou, M., and Sasikala, K.R., The Depster-Shafer Combination Rule as a Tool to Classifier Combination. Geoscience and Remote Sensing Symposium. Proc. IGARSS, 2000, IEEE Press, pp. 2429-2431.

[9]Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, Proc. of the 14^{th} Int. Joint Conf. on A. I.., Vol. 2, Canada, 1995, pp. 1137-1143.

[10] Carletta, J.. Assessing Agreement on Classification Taks: the Kappa Statistic, Computational Linguistics, v. 22(2), 1996, pp. 249-254.

[11] Congalton, R. G., A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data, Remote Sensing Environment, v.37, 1991, pp. 35-46.

[12] Cohen, J., A Coefficient of Agreement for Nominal Scales, Educational and Psychological Measurement. v.20, n.1, 1960, p. 37-46.

[13] Duda, R. O., Hart, P.E. and Stork, D. G. "Pattern Classification", *Wiley*, 2ed, New York, 2000.

[14] Nguyen, D. and Widrow, B. Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights," International Joint Conference of Neural Networks, v. 3, pp. 21 - 26, 1990.