# Fuzzy classification of spectra

Theo E. Schouten and Maurice S. klein Gebbinck

Katholieke Universiteit Nijmegen, Computing Science Institute
Toernooiveld 1, 6525 ED Nijmegen The Netherlands
Email: ths@cs.kun.nl

**Abstract.** Measured spectra are often a composition of a number of basic spectra. The purpose of fuzzy classification is to determine the presence and abundance of the basic spectra in a measured spectrum. In this paper spectra from remote sensing imaging spectrometers with 6 resp. 220 spectral bands are considered each composed of 3 basic spectra from a set of 6 resp. 11 basic spectra. It will be shown that a backpropagation neural network is able to determine the types and abundances of the basic spectra in a mixed spectrum.

## 1 Introduction

Measured spectra are often a composition of a number of basic spectra. In remote sensing, imaging spectrometers [1] acquire images simultaneously in many narrow spectral bands from visible to infrared. Due to the limited spatial resolution, a pixel often contains more than one ground cover e.g. grass, mais and a water filled ditch in between. The purpose of fuzzy classification or mixed pixel analysis is to determine the fractions of the ground cover categories (called endmembers) present in each pixel.

Usually a linear mixture model [2] is used, the signal received for a pixel in band $i$ is assumed to be: $s_i = \sum_{j=1}^{c} f_j M_{ij} + e_i, \quad i = 1, 2, ..., n$. Here $n$ is the total number of bands, $c$ is the total number of end-members in the pixel, $M_{ij}$ is the reflectance of the $j$th end-member in the $i$th band, $f_j$ is the fraction of the pixel covered by end-member $j$ with the constraints: $\sum_j^c f_j = 1$ and $0 \leq f_j$ for all $j$, and $e_i$ is the error in the $i$th band.

In classical methods [3, 4] the error is assumed to have a multivariate normal distribution with zero mean and covariance matrix $N_f$ determined by the covariance matrices of the endmembers: $N_f = \sum_{j=1}^{c} f_j N_j$. Following the maximum likelihood approach the $f$ is selected which minimizes: $e^T N_f^{-1} e + ln|N_f|$. When all $N_j$ are equal to say $N$ analytical solutions can be derived. Ignoring the constraints on $f$ gives: $\hat{f}_0 = UM^T N^{-1} s \;\; with \;\; U^{-1} = M^T N^{-1} M$. Imposing the sum-to-one constraint gives: $\hat{f}_1 = \hat{f}_0 + \alpha(1 - I^T \hat{f}_0)UI \;\; with \; \alpha = (I^T UI)^{-1}$, $I$ is the identity vector. The positivity constraint is then usually taken into account by setting negative $f_j$'s to zero and then renormalising the $f$. To obtain more accurate estimations of $f$ numerical methods must be used involving searches through the $f$ space, which are very time consuming.

In this study a back-propagation neural network is used to perform the mapping from $s$ to $f$. The used images, the generation of the training and testing data sets, and the neural network used are described in section 2. In the last section the obtained results and our conclusions are given.

## 2 Data set generation

| Landsat (6 bands) | | | Aviris (220 bands) | | |
|---|---|---|---|---|---|
| Endmember | Learn | Test | Endmember | Learn | Test |
| Soil | 475 | 240 | Bldg-Drives | 66 | 36 |
| Water | 165 | 79 | Woods | 740 | 361 |
| Vineyard | 529 | 264 | Corn-notill | 311 | 149 |
| Weeds | 233 | 107 | Corn-min | 83 | 52 |
| Garigue | 663 | 324 | Grass/Pasture | 184 | 89 |
| Grass | 1579 | 785 | Grass/Trees | 369 | 178 |
| | | | Hay-windrowed | 312 | 152 |
| | | | Soybeans-notill | 176 | 81 |
| | | | Soybeans-min | 680 | 352 |
| | | | Soybean-clean | 58 | 27 |
| | | | Wheat | 128 | 62 |

**Table 1.** Number of pixels in learn and test set per class per spectrometer

Pixel spectra from two spectrometers were available, the class of each pixel being determined by ground survey or classification of an image. From a 6 band Landsat image of the Lisbon area in Portugal [5] 6 classes were taken. From the 220 band AVIRIS data accompanying the MultiSpec computer program [6] 11 classes were selected. In table 1 the classes and the number of pixels in a learn and a test set are given.

From the pure pixels the mixed spectra are generated in the following way. All combinations of 3 classes of which 1 belongs to the first 2 classes (considered to be a boundary between 2 agricultural fields) are considered. For each combination the $f$ space is covered by a grid with a certain spacing (usually set to 0.025) and at each point a number (usually 2 or 10) of mixed spectra are calculated as: $s_i = \sum_{j=1}^{c} f_j M_{ij}$ with $M$ randomly selected from the pure pixels. Train and test sets are so generated for both spectrometers.

As a neural network the multi-layer back-propagation [7] net was used. The simulations were performed with our network simulator program, which executes on a PowerXplorer parallel machine and is implemented in C using PVM for communication. Based on our experiences with other applications the following strategies were used. The input and output values were scaled linearly to the range -1.0 to 1.0. The transfer functions were taken as $tanh$ for the hidden layers and linear for the output layer. The learning rate was varied by hand during learning, starting with a high value such that no oscillations in the error occur, and decreasing it when the decrease in error becomes too low. The momentum rate was typically set to 0.8.

Learning was done in epoch mode but weights and biases were updated several times (typically 50) during each epoch. This semi-epoch mode approximates on-line learning which in our experience usually produces somewhat better results than full epoch learning when there are a large number of training samples. On a sequential machine on-line learning is usually faster (that is: the weights with best results on the test data are quicker obtained) than epoch learning. On a parallel machine however different parallelization strategies[8] can be used for on-line and (semi-)epoch learning, making the latter faster as less communication operations are needed. The above quality and learning time observations are off course dependent on the type of problem, the structure of its error surface, the size of the network, the number of training patterns and the ratio between communication and calculation speed of the parallel machine used.

The number of hidden layers and neurons appeared to be not very critical, comparable results were obtained over a wide range. Finally a network with one hidden layer was chosen with 64 neurons for the Landsat and 32 neurons for AVIRIS data.

## 3 Results and conclusions

| | Neural Network | | Analytical | | Numerical | |
|---|---|---|---|---|---|---|
| $f_{j\ true}$ | Average | $\sigma$ | Average | $\sigma$ | Average | $\sigma$ |
| 0.00 - 0.05 | 0.05 | 0.07 | 0.08 | 0.13 | 0.05 | 0.09 |
| 0.05 - 0.15 | 0.00 | 0.07 | 0.04 | 0.14 | 0.00 | 0.09 |
| 0.15 - 0.25 | -0.05 | 0.08 | -0.03 | 0.16 | -0.04 | 0.11 |
| 0.25 - 0.35 | -0.06 | 0.10 | -0.07 | 0.17 | -0.06 | 0.13 |
| 0.35 - 0.45 | -0.07 | 0.11 | -0.11 | 0.18 | -0.07 | 0.14 |
| 0.45 - 0.55 | -0.07 | 0.12 | -0.15 | 0.19 | -0.08 | 0.15 |
| 0.55 - 0.65 | -0.07 | 0.12 | -0.19 | 0.20 | -0.09 | 0.16 |
| 0.65 - 0.75 | -0.08 | 0.13 | -0.24 | 0.21 | -0.10 | 0.17 |
| 0.75 - 0.85 | -0.08 | 0.12 | -0.29 | 0.22 | -0.11 | 0.18 |
| 0.85 - 0.95 | -0.10 | 0.09 | -0.37 | 0.23 | -0.13 | 0.19 |
| 0.95 - 1.00 | -0.13 | 0.09 | -0.43 | 0.23 | -0.16 | 0.19 |

**Table 2.** Landsat, Average and $\sigma$ of $f_{j\ calculated} - f_{j\ true}$

In tables 2 and 3 the obtained results on the test sets are given. The analytical method gives the worst results, not useful for many applications but it is a fast method. The classical numerical method gives better results, but it is slow: 3 days on a 16 node PowerXplorer for AVIRIS for 130000 pixels. The neural network gives the best results for the Landsat data and is comparable to the numerical method for the AVIRIS data. Its speed is comparable to the analytical method.

| $f_{j\ true}$ | Neural Network Average | Neural Network $\sigma$ | Analytical Average | Analytical $\sigma$ | Numerical Average | Numerical $\sigma$ |
|---|---|---|---|---|---|---|
| 0.00 - 0.05 | 0.03 | 0.05 | 0.05 | 0.08 | 0.03 | 0.05 |
| 0.05 - 0.15 | -0.02 | 0.06 | -0.02 | 0.09 | -0.03 | 0.06 |
| 0.15 - 0.25 | -0.07 | 0.07 | -0.07 | 0.10 | -0.07 | 0.09 |
| 0.25 - 0.35 | -0.10 | 0.09 | -0.12 | 0.11 | -0.10 | 0.11 |
| 0.35 - 0.45 | -0.12 | 0.11 | -0.16 | 0.11 | -0.11 | 0.12 |
| 0.45 - 0.55 | -0.12 | 0.12 | -0.21 | 0.12 | -0.10 | 0.12 |
| 0.55 - 0.65 | -0.13 | 0.13 | -0.25 | 0.13 | -0.10 | 0.12 |
| 0.65 - 0.75 | -0.14 | 0.13 | -0.30 | 0.14 | -0.12 | 0.13 |
| 0.75 - 0.85 | -0.16 | 0.12 | -0.36 | 0.15 | -0.13 | 0.12 |
| 0.85 - 0.95 | -0.22 | 0.12 | -0.45 | 0.15 | -0.16 | 0.10 |
| 0.95 - 1.00 | -0.28 | 0.10 | -0.41 | 0.14 | -0.13 | 0.08 |

**Table 3.** AVIRIS, Average and $\sigma$ of $f_{j\ calculated} - f_{j\ true}$

A back-propagation neural network with one hidden layer is thus able to learn the relation between the intensities of a pixel and the fractions of its ground cover categories. This was demonstrated on data from 2 spectrometers with 6 resp. 220 bands using 6 resp. 11 end-members.

Training of the neural networks requires the construction of training and test data sets and is relative straightforward but very time consuming, in the order of days.

## References


1. G. Vane et al: The Airborne Visible/Infrared Imaging Spectrometer(AVIRIS). Remote Sensing of Environment **44** (1993) 127–143
2. H.M. Horwitz et al.: Estimating the proportion of objects within a single resolution element of a multispectral scanner. Proceedings of the 7th International Symposium on Remote Sensing of Environment, Ann Harbor, Michigan (1971) 1307-1320
3. J.J. Settle, N.A. Drake: Linear mixing and the estimation of ground cover proportions. International Journal of Remote Sensing **14** (1993) 1159–1177
4. M.S. klein Gebbinck, T.E. Schouten: Decomposition of mixed pixels. Image and Signal Processing for Remote Sensing II Proc. SPIE 2579 (1995)
5. R.P.H.M. Schoenmakers, G.G. Wilkinson, T.E.Schouten: Results of a hybrid segmentation method. Image and Signal Processing for Remote Sensing I Proc. SPIE 2315 (1994) 113-126
6. D. Landgrebe: Multispectral Data Analysis: A Signal Theory Perspective. Purdue University June 1996, `http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/`
7. D.E. Rumelhart, G.E. Hinton, R.J. Williams (Eds.): Learning internal representations by error propagation. Parallel Distributed Processing (1996) chap. 8, 318–362
8. Louis Vuurpijl, Theo Schouten, Jan Vytopil: Performance Prediction of Large MIMD Systems for Parallel Neural Network Simulations. Future Generation Computer Systems **11** (1995) 221–232


This article was processed using the LaTeX macro package with LLNCS style