REDUCTION OF THE DIMENSIONALITY OF HYPERSPECTRAL DATA FOR THE CLASSIFICATION OF AGRICULTURAL SCENES

Claudionor Ribeiro da SILVA[1], Jorge Antônio Silva CENTENO[1] and Selma Regina Aranha RIBEIRO[2]

[1] *Federal University of Paraná, Brazil (crs and centeno)@ufpr.br*

[2]*State University of Ponta Grossa, Brazil (selmar@ufpr.br)*

Abstract: Recent advances in sensor technology opened new possibilities for remote sensing. For example, the appearance of sensor higher spatial and spectral resolution. In terms of spectral resolution, the number of available bands increased significantly, resulting in hyperspectral sensors. Hyperspectral remote sensing images are characterized by the division of the electromagnetic spectrum in a great number of narrow spectral bands, which enable greater detail of spectral variation of targets. High dimensionality demands special attention in the classification process. The main problem caused by the increase of the dimensionality is the reduction of the efficiency of the classifiers. This problem is known as the Hughes phenomenon. The occurrence of the Hughes phenomenon is caused by the exaggerated increase of the dimensions of the variance covariance matrix (increase of the dimensionality), compared to the limited number of available training samples. As a result, recent approaches focus reduction of the dimensionality. In this paper, a method of feature selection from hyperspectral images is presented. The proposed method, based in the use of the Genetic Algorithms, is evaluated with a AVIRIS data set and the results are compared to the results of other algorithms (Sequential Forward Selection and Sequential Backward Selection), recognized as techniques for reduction of dimensionality. A Genetic Algorithm can be described as a global search technique for optimization purposes inspired by the natural evolutionary process. The experiments show that Genetic Algorithms based reduction method can be used to reduce the dimensionality within image classification in remote sensing.

1. INTRODUCTION

In remote sensing, the number of available bands has increased significantly, allowing the use of hyperspectral images. This allows for obtaining greater detail of the spectral variation of a target. One of the main problems caused by the increase in dimensionality is the Hughes phenomenon, which causes a reduction of the efficiency of digital classifiers. This problem is caused is caused by the increase increase of the dimensionality of the covariance matrix, compared to the number of available training samples. Some solutions have been proposed to address this problem, such as the regularization of the covariance matrix (Friedman, 1989) or the reduction of dimensionality (Bruske and Merényi, 1999; Serpico, 2002; Du and Qi, 2004; Zortea et al., 2005). on Genetic Algorithms to select the best suited spectral bands for

classification purposes. There are two main reasons for the reduction of the dimensionality of hyperspecral data. First, a great computational effort is necessary to process the data. The second reason is because a high dimensionality reduces the accuracy of the classification, as discussed above. A selection method based on genetic algorithms was proposed, assuming that the selected variables will be used for classification purposes. In the next section, a brief description of the methods and concepts used in this paper are presented.

2. BIBLIOGRAPHIC REVIEW ON REDUCTION OF DIMENSIONALITY

One of the most popular methods for feature reduction in remote sensing is the use of the principal components transform (Du and Qi, 2004; Wang and Chang, 2006). The Principal Components PC transformation transforms the original data into a new smaller set, which are less correlated as the first dataset. Therefore, a reduced number of new variables may content the information of the original set. However, although it is frequently used, the PC transform is not a appropriate for feature extraction in the classification, because it is does not consider the classes of interest, but only the data set. Therefore, it may not produce the optimum subspace for the classification. Others methods to reduce dimensionality are shows in (Bruske and Merényi, 1999; Zhang et al. 1999; Lennon, et al. 2001; Lin and Bruce, 2003; Bittencourt and Clarke, 2004).

Instead of computing new variables from the original set, Feature Selection algorithms select the most relevant variables from the original set. The main algorithms, tested in Bon (2001), are Sequential Forward Selection – SFS and Sequential Backward Selection – SBS. Those algorithms have the property of conserving a group of variables that better represent the main characteristics of a set. The first algorithm (SFS) eliminates poor variables from the original set, preserving a sub-set with the best variables. On the other hand, the second one (SBS) selects the best variables from the original set, creating a sub-set with the best variables. If the discriminate function of the two algorithms is the same, the result is similar (Bon, 2001).

The Sequential Forward Selection is a bottom-up strategy that selects the most significant variables using an iterative process (Bon, 2001; Kumar et al., 2001). The algorithm is described by Bon (2001). Let X be a set of original variables, for example a set of spectral bands; d the number of the original variables, Y a set of selected variables, and Z the set of remaining variables after the selection of $\mathrm{Y}\,(\mathrm{Z} = \mathrm{X} - \mathrm{Y})$. At the beginning, the set of optimal variables is empty $Y_0 = \phi$ and the set Z is equal to X. A function $\mathrm{J}()$ that identifies the most significant variables within the set Z also needs to be defined before an iterative process is started. In each iteration, the most significant variable is chosen from the remaining set (Z). After k iterations, k selected variables form a new set $\mathrm{Y_k}$. In order to select the most significant variable, all available variables of set $\mathrm{Z_k}$ are ordered according to the value of the function $\mathrm{J}()$ and the variable with the highest value is chosen and moved to the set $\mathrm{Y_k}$, building the new set $\mathrm{Y_{k+1}}$. In the algorithms SFS the Bhattacharya distance was used as a criterion of select the variables ($\mathrm{J}()$). According to Serpico et al. (2002) this method is computationally very efficient. A disadvantage of the method is the fact that, after a variable is chosen, the decision can not be changed in the next iterations.

The approach of the Sequential Backward Selection SBS method is based on a top-down strategy. It is also an iterative process, but the difference is that less significant variables are discarded from the original set (Bon 2001) within an iterative process. A function J() to evaluate the significance of the variables is also necessary. At the beginning, the set of

selected variables is equal to the original set ($Y_0 = X$). In each iteration the less significant variable is identified and eliminated from the set Y_k. The new set ($Y_{k+1} = Z$) has fewer elements and is used in the next iteration. The process is repeated, until the desired amount of selected variables set is reached. In the algorithms SBS the Jeffries Matusita distance was used as a criterion of select the variables (J()).

3. CLASSIFICATION

For the experiments, image classification was performed using the Gaussian Maximum Likelihood classifier method. A description of the method can be found in Richards and Jia (1998) and can be described as follows:

Pixel x, belongs to class wi if $P(x \in w_i) > P(x \in w_j)\ \ i \neq j$:

$$P(w_i \mid x) > P(w_j \mid x) \quad with \quad P(w_i \mid x) = \frac{P(x \mid w_i) P(w_i)}{P(x)} \tag{1}$$

Assuming a Gaussian distribution for the pixels within a class, an estimate of $P(x \mid w)$ can be obtained from the training set:

$$P(x \mid w_i) = (2\pi)^{-\frac{n}{2}} \left|\sum_j\right|^{-\frac{1}{2}} \exp(-0.5(x - \mu_j)^t \sum^{-1} (x - \mu_j)) \tag{2}$$

here, $\sum_j$ stands for the variance covariance matrix, $\left|\sum_j\right|$ stands for the determinant of the variance covariance matrix of class j, and μ_j for its mean vector.

After the classification step, the accuracy of the resulting thematic image has computed. A value that describes the accuracy of the thematic map can be obtained from the confusion matrix, for example the global accuracy, the producer's accuracy, the user's accuracy or the kappa index. The kappa index is computed from the confusion matrix as follows (Richards and Jia 1998):

$$k = \frac{n \sum_{i=1}^{M} D(i,i) - \sum_{i=1}^{M} SL(i) SC(i)}{n^2 - \sum_{i=1}^{M} SL(i) SC(i)} \tag{3}$$

where n stands for the number of spectral bands; $D(i,i)$ stands for the element of the variance covariance matrix at row i, column j, SL stands for the sum of the elements of the same row and SC the sum of the elements of a given column.

4. FEATURE REDUCTION WITH GENETIC ALGORITHMS

The method is based on the genetic algorithms (GA) proposed by Goldberg (1989), that are a global search technique for optimization purposes inspired by the Natural Selection process. Its operators resemble biological evolutionary processes, such as crossover, selection and mutation. The aim of a genetic algorithm is to find the optimal solution for a problem, combining suboptimal solutions within a process that converges to the optimal one. Because the optimal solution is a priori unknown, the process starts with an arbitrary set of solutions as its initial population. In order to adapt the algorithm for a given problem, the solution must be

coded, building a binary chain that resembles the biological genetic chain. Within an iterative process, the best solutions are identified and combined, generating a new set of better solutions. So, the worst solutions are discarded and the population is optimized. This process is repeated until an optimal solution is reached, which means that the solution of the problem was found, or it is stopped using an external criterion, like a fixed number of iterations, obtaining a sub-optimal solution.

4.1. Binary Coding

In our problem, the binary chain has a size equal to the number of original variables (large bands). Each element of the chain has two possible statuses: 0 if the spectral band is not suitable for the classification and should be discarded and, 1 if the spectral band is suitable for the classification and should be preserved. After selecting the coding option, many solutions are randomly generated, building an initial population.

4.2. Selection

During each successive step, a small set of solutions is selected from the current population. Therefore, the fitness of each solution is evaluated and the worst solutions, according to its fitness, are discarded. The fitness function must describe the suitability of the solution for the given problem and must be proposed by the user. Optimal chromosomes, chromosomes which the best fitness value, are used to build a new population by breeding (crossover), which is done combining their genetic chains. The fitness function presents a solution for the pointed problem, being crucial for the algorithm.

4.3. Reproduction

After computing the fitness for each solution, the most suitable elements are selected and used to generate the next generation, applying genetic operators, like crossover and mutation Goldberg (1989). Crossover has the purpose to generate new individuals (solutions) from two or more members of the population (parents), combining its respective binary chains. Mutation is used to introduce new genetic material into the process making it easier to reach a global minimum within the search region.

4.3.1. Crossover

The simple crossover operator, combining the chains of two parents, was used. First, a crossover point is selected randomly. Then, the chains of the parents are broken at the crossover point and two offspring are generated. One receives the first part of the chain of the first parent and the second part of the chain of the second parent. The other offspring receives the remaining parts of the first and second parent.

4.3.2. Mutation

Mutation consists of randomly inverting the value of a bit of the binary chain. Hence, random modifications are introduced in the individuals, as it is shown in figure 3 (b). This operator is essential for accelerating the search process. When the algorithm meets a local minimum, the mutation provokes variations in the genetic material that excite the search, focusing on other regions within the search area towards a global minimum.

4.4. End of the Process

This process is repeated until a termination condition is reached. The condition can be set based on criteria like reaching a minimum fitness value, reaching a maximal number of iterations or achieving the desired number of bands. In this experiment the last criterion was used, where the required number of bands is fixed by the user, before beginning the iterations. That criterion was chosen because the interest exists in a certain number of bands.

4.5. Fitness Function for Band Selection

We propose a fitness function for the feature selection problem that is based on three main goals:

1. to reduce the number of features,
2. to preserve spectral separability between classes, and
3. to discard highly correlated bands.

The fitness function is the mean of the following three factors:

1. A(DT): a factor related to the separability, derived from the Transformed Divergence: The aim is to preserve spectral separability between classes, using the transformed divergence (Swain and Davis, 1978). The value of the divergence was computed for each pair of classes and the minimum value was used to compute a value of the transformed divergence that will represent the bands set.
2. $A(C_{i,j})$: the correlation factor: The factor is related to the correlation between bands. The values of the correlation ($C_{i,j}$) between each pair of bands within the current set were computed. Because high correlation values are unwanted, the maximum correlation value was used to compute the factor.
3. A(NR): a factor related to the number of spectral bands of the resulting set: The minimum number of bands that represent the original set without considerable loss of information was determined by the method of Cadima (2001). For the set of bands used in this experiment, the computed minimum value is equal to 1,516, so the minimum number of bands was set to 2. The factor associated to the number of bans grows as the number of bands increases.

So, the fitness function can be described as follows:

$$fitness = aA(DT) + bA(C_{i,j}) + cA(NR) \quad (3)$$

where a, b and c are constant belonging to the interval $[0,1]$.

Once the fitness function is defined, it can be applied to rank all individuals according to their fitness and find the best and worst solutions. To guarantee that the best solutions of the previous iteration are preserved, a small parcel of the individuals, with higher fitness, can be used to compose part of the new population. This is called elitism. The number of preserved individuals is defined by the user. A high rate of elitism in not optimal, because it slows in the process convergence or, even, make it difficult to converge on an optimal solution.

After the elitism step, the population is completed by combining pairs of individuals of the previous iteration. Two of the selection techniques, introduced by Goldberg (1989), were used referred to as tournament and roulette. Tournament can be defined as follows: a pair of individuals is randomly chosen from the population. Comparing their fitness values, the individual with lower fitness is discarded and the best preserved. This is repeated until enough

new solutions, to compose a new population, are generated. When using the roulette method, a pair of individuals is chosen randomly from the original set.

5. EXPERIMENTS

A hyperspectral reflectance image (figure 1), covering a test site, in Indian Pines, in Northwestern Indiana/USA, was used to perform the tests. Although AVIRIS sensor to own 224 bands, the image supplied by Purdue it has 220 bands. The dataset acquired by the AVIRIS sensor in 1994, has a spatial resolution of 20m, and was distributed by search. The necessary ground truth was obtained by the research group of the Purdue University, IN. A small region, with 70x70 pixels was used to carry out the tests (White square in the figure 1). Four classes were chosen to perform the classification: corn, soy, vegetation (trees) and soy minimum cultivation (initial phase of the culture). The region was chosen because a detailed field survey was carried out as the image was recorded by AVIRIS.
Although the AVIRIS dataset is composed by 220 narrow spectral bands, some of them are not useful, because of atmospheric water absorption. In the experiment, these bands were included so as to evaluate the capacity of the genetic algorithms to distinguish noise from useful information.

Figure 1 - Hyperspectral image - band 15 (0,54µm) (Indiana/USA, 1994)

For the experiment, taking into account the number of spectral bands, a 220 elements chain was used. The aim of the experiment was to select the best set of spectral bands to classify the image into the four classes discussed above. The procedure was started with a population of 20 individuals. The size of the population is chosen empirically, through preliminary experiments. Selection by tournament, elitism, crossover and mutation were used in the evolutionary process.
Using the same training areas, the SFS and SBS algorithms were used to reduce the dimensionality of the same variable set. In the last step, digital classification was performed using the solutions obtained applying the three methods. The accuracy of the results was evaluated using kappa index and the global accuracy.

6. RESULTS AND DISCUSION

Due to high spectral correlation among bands neighboring different results can be obtained, by genetic algorithms, in the process of dimensionality reduction. The original dataset includes images that are strongly affected by noise. The Genetic Algorithms, in all cases, are able to discard such noisy bands. The best result obtained using Genetic Algorithms will be here discussed. The distribution of the selected spectral bands along the spectrum depends on the training samples that describe the classes. In the experiment, the classes (corn, soy, vegetation and soy minimum cultivation) are spectral mixtures of soil and vegetation, therefore, the concentration of bands in the SWIR was expected. In spite of the concentration of bands in SWIR, the bands selected in the visible and in the near infrared also have important function in the classification process.

A comparative study of the genetic algorithms and the sequential algorithms (SFS and SBS), in terms of processing time, was performed, using a 2.2. Sempron computer, with 256 Gb RAM. The algorithms were coded using MATLAB 5.3. Three groups were chosen, arbitrary, with 15, 30 and 45 bands, to show a summary of processing time (table 1). The sequential algorithms are characterized by a high speed. This fact can be attributed to its easy implementation. Comparing the genetic algorithms, the elitism based algorithm worked more efficiently. Using only elitism demands fewer operations than the tournament approach.

Groups	Number of bands	Sequential algorithms		Genetic algorithms	
		SFS	SBS	tournament	elitism
5	15	1,219	1,641	3,579	2,778
10	30	1,281	1,656	3,678	3,013
15	45	1,297	1,696	3,981	3,268

Table 1- Processing time (seconds)

Using the set with 21 spectral bands obtained in each method, the image was classified and thematic maps were produced. Table 2 presents the values of the kappa index for each method. The reliability of the kappa index values was evaluated by means of statistical inference. All the values lie inside of a 95% confidence interval.

Algorithms	Kappa Index
AG's - Elitism	0,9218
AG's - Tournament	0,9067
SFS	0,8497
SBS	0,8317

Table 2 - Better indexes kappa

7. CONCLUSION

The experiments performed have proved the feasibility of the use of genetic algorithms to reduce the dimensionality of hyperspectral remote sensing data to improve the accuracy of digital classification. The best results were obtained with the elitism-based algorithm. One advantage of the use of a genetic algorithm is that a spectral band that is discarded within the evolutional process can be reintroduced in the optimal solution through reproduction or mutation, which allows more flexibility in the search for an optimal solution. The genetic

algorithm was also able to identify and discard noisy bands, based on the fitness criterion computed from the correlation, transformed divergence and optimal number of bands.
In future experiments the evolutionary process and the function fitness used into genetic algorithms will be revalued. Attention will be given to other selection techniques and measures of spectral similarly.

References

Bittencourt, H. R.; Clarke, R. T. (2004). *Feature Selection by Using Classification and Regression Trees (CART)*. International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences. Natural Resources Canadá. v. 35, ISSN 1682-1750, p.66-70.

Bon, F. J. C. (2005). *Selección y Extracción de Características*. available at: http://www.etsi2.ugr.es/depar/ccia/rf/www/tema5_00-01_www/node/node16.html. Access date: 19 dec.

Botelho, S. C.; Bem, R. de; Figueiredo, M.; Lautenschlger, W.; Centeno, T.; Mata, M. (2005). *C-NLPCA: Extracting Non-Linear Principal Components of Image Datasets*. XII Simpósio Brasileiro de Sensoriamento Remoto. Anais do SBSR. Goiânia, p.3495-3502.

Bruske, J.; Merényi, E. (1999). *Estimating the Intrinsic Dimensionality of Hyperspectral Images*. 1999. ESANN'1999 proceedings - European Symposium on Artificial Neural Networks Bruges; ISBN 2-600049-9-X. Belgium, p.105-110.

Cadima, J. F. C. L. (2001). *Redução de Dimensionalidade Através Duma Análise em Componentes Principais: Um Critério para o Número de Componentes Principais a Reter.* Revista de Estatística (INE), Portugal, p. 37-49.

Chuvieco, E. (1990). *Fundamentos de Teledetección Espacial*. Madrid, Ediciones Rialp/AS.

Du, H.; Qi, H. (2004). *An FPGA Implementation of Parallel ICA for Dimensionality Reduction in Hyperspectral Images.* Geoscience and Remote Sensing Symposium (IGARSS), Proceedings. IEEE International. v.5, p.3257-3260.

Friedman, J. H. (1989). *Regularized Discriminant Analysis*. Journal of American Statistical, v. 84, n. 405. p.165-175.

Gallo, I.; Binaghi, E.; Boschetti, M.; Brivio, P. A. (2004). *A Neural Adaptive Model for feature selection and hyperspectral data classification.* Image and Signal Processing for Remote Sensing IX. v. 5573, ISSN 0277-786X.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. New York, Addison-Wesley.

Hughes, G. F. (1968). *On the Mean Accuracy of Statistical Pattern Recognizers*. IEEE - Transactions on Information Theory, v. IT-14, n. 1.

Kumar, S.; Ghosh, J.; Crawford, M. M. (2001). *Best-Bases Feature Extraction Algorithms for Classification of Hyperspectral Data*. IEEE - Transactions on Geoscience and Remote Sensing; v. 39, n. 7.

Lennon, M.; Mercier, G.; Mouchot, M. C.; Hubert-Moy, L. (2001). *Curvilinear Component Analysis for Nonlinear Dimensionality Reduction of Hyperspectral Images*. Proceedings. SPIE the International Society for Optical Engineering - International Society for Optical Engineering. p.157-168.

Lin, H.; Bruce, L. M. (2003). *Parametric Projection Pursuit for Dimensionality Reduction of Hyperspectral Data. Proceedings.* IEEE International Geoscience and Remote Sensing Symposium (IGARSS). v. 6, p.3483-3485.

Mitchell, M. (1997). *An Introduction to Genetic Algorithms*. Cambridge, MIT Press.

Moigne, J. L. (2005). *Multi-Sensor Image Registration, Fusion and Dimension Reduction*. Available at: http://www.esto.nasa.gov/conferences. Access date: 31 aug.

Pla, L. E. (1986). *Analisis Multivariado: Metodo de Componentes Principales*. Universidad Nacional Experimental Francisco de Miranda Coro. Falcón, Venezuela, p.29-38.

Richards, J. A.; Jia, X. (1998). *Remote Sensing Digital Image Analysis – an Introduction*. Australia, Third edition, Springer.

Salehi, B.; Zouj M.J.V. (2004). *Wavele-Based Reduction of Hyperspectral Imagery*. XXth ISPRS Congress. Istanbul/Turkey.

Serpico, S. B.; D'inca, M.; Melgani, F.; Moser, G. (2002). *A comparison of feature reduction techniques for classification of hyperspectral remote-sensing data*. Image and Signal Processing of Remote Sensing VIII. v. 4885.

Swain, P. H.; Davis, M. S. (1978). *Remote Sensing the Quantitative Approach*. New York, McGraw-Hill.

Zeng, H.; Trussell, H.J. (2004). *Dimensionality reduction in hyperspectral image classification*. IEEE - International Conference on Image Processing. v. 2, p.913-916.

Zhang, Y.; Desai, M. D.; Zhang, J.; Jin, M. (1999). *Adaptive Subspace Decomposition for Hyperspectral Data Dimensionality Reduction.* IEEE - International Conference on Image Processing. v. 2, p.326-329.

Zortea, M.; Haertel, V.; Clarke, R. T. (2005). *Investigação Sobre Métodos para Redução da Dimensionalidade dos Dados em Imagens Hiperespectrais*. Revista Brasileira de Cartografia, Nº 57/01, ISSN 0560-4613, p.36-42.

Wang, J.; Chang, C. (2006). *Independent Component Analysis-Based Dimensionality Reduction With Applications in Hyperspectral Image Analysis*. IEEE - Transactions on Geoscience and Remote Sensing; v. 44, n. 6.

Corresponding author contacts

Claudionor Ribeiro da SILVA
crs@ufpr.br
Federal University of Paraná
Brazil