Anais XIII Simpósio Brasileiro de Sensoriamento Remoto, Florianópolis, Brasil, 21-26 abril 2007, INPE, p. 6431-6436.

Single and multiple stage classifiers implementing logistic discrimination

Hélio Radke Bittencourt [1]
Denis Altieri de Oliveira Moraes [2]
Victor Haertel [2]

[1] Pontifícia Universidade Católica do Rio Grande do Sul - PUCRS
Av. Ipiranga, 6681, CEP 90619-900, Porto Alegre, RS, Brazil
heliorb@pucrs.br

[2] Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, UFRGS
Av. Bento Gonçalves, 9500, CEP 91501-970, RS, Brazil
{daltieri, victor.haertel}@ufrgs.br

Abstract. The logistic discrimination technique can be regarded as a partially parametric approach to pattern recognition. This method is general and robust because it doesn't make assumptions on the underlying distribution of data. Moreover, it requires the estimation of fewer parameters than some better-known procedures such as the Gaussian maximum likelihood discriminator. In this paper, two different approaches to implement logistic discrimination are presented. Experiments were performed using high-dimensional image data acquired by the AVIRIS system, showing a test area which includes a number of classes spectrally very similar. Results are presented and discussed

Keywords: remote sensing, image processing, logistic discrimination, sensoriamento remoto, processamento de imagens, discriminação logística.

1. Introduction

Statistical pattern recognition applied to digital image classification has been extensively discussed in the scientific literature. According to Jain et al (2000), from 1979 to 2000, 350 articles dealing with pattern recognition were published in IEEE Transactions on Pattern Analysis and Machine Intelligence. Out of this total, 85% were related to statistical pattern recognition. In the statistical approach, each pattern is regarded as a p-dimensional random vector, where p is the number of variables used in classification. A number of statistical methods for classification have been proposed, each one presenting advantages and disadvantages.

According Bittencourt and Clarke (2003) techniques requiring assumptions on the functional shape of the variables in the feature space involve parameter estimation, and are therefore termed parametric. Depending on the classification method, the number of classes present and the number of variables used, the number of parameters to be estimated may vary substantially. In the Gaussian maximum-likelihood method, par example, the underlying probability distributions are assumed to be multivariate Normal and the number of parameters to be estimated can be very large. Haertel and Landgrebe (1999) states that that the estimation of the within-class covariance matrices is one of the most difficult problems in dealing with high-dimensional data, especially due the fact that the number of available training samples is frequently very limited.

Methods based on logistic discrimination have advantages compared to other parametric methods because the assumptions required are considerably weaker and the number of parameters to be estimated is smaller. In this paper two possible approaches to pattern classification using based on logistic model are investigated. The first one – single stage – implements the concepts developed in logistic discrimination, as derived from the multinomial logistic regression model. The second one – multiple stage – consists in successive applications of a traditional binary logistic model, using a distance measure to select the classes to be dealt with at each stage.

In the following sections these two logistic models are illustrated using AVIRIS hyperspectral image data. The results are presented and discussed.

2. Logistic Discrimination

Logistic Discrimination was originally proposed by Bittencourt and Clarke (2000) in the classification of remote sensing image data of natural scenes. The main advantage of this approach as compared with the more traditional quadratic classifiers such as the Gaussian maximum likelihood lies in the smaller number of parameters to be estimated. In this case, the probability that a pattern x belongs to the class wi can be directly estimated by

$$P(w_i \mid \mathrm{x}) = \frac{\exp\left(\beta_{i0} + \beta_i^T \mathrm{x}\right)}{1 + \sum_{j=1}^{k-1} \exp\left(\beta_{j0} + \beta_j^T \mathrm{x}\right)}$$

Here, β_{i0} and $\boldsymbol{\beta}_i$ are the model parameters, with β_{i0} termed the intercept and $\boldsymbol{\beta}_i$ is a vector of parameters associated with the *p* characteristics of the vector **x**. The logistic model requires the estimation of *k-1* vectors involving the parameters $\boldsymbol{\beta}_i$, corresponding to the *k-1* classes present in the image. The *k-th* class is taken as a basis, from which the natural log of the ratio of the two probabilities become linear functions of the parameters. This logarithm is known as the logit function. The reduction in the number of parameters plays an important role whenever the number of training samples available is small compared to the data dimensionality. Therefore, the logistic discrimination may be of particular interest in remote sensing hyperspectral image data classification. Another possible approach to deal with the small sample size problem is the implementation of the multiple stage approach in the classification process. In this case, several traditional binary logistic regressions are used at each stage, each one dealing with a pair of classes at a time. Decision tree classifier is a type of hierarchical classifier that has been investigated by several authors such as Moraes (2005) and Bittencourt and Clarke (2003b).

3. Single Stage and Multiple Stage Classifiers

An example of the traditional single stage classifier methodology applied to a four classes problem is illustrated in **Figure 1**. The patterns to be labeled (pixels) are initially clustered in a single group. Decision functions are then estimated for each individual class under consideration. Next, each individual pattern is inserted into the decisions functions and labeled according to the winner.

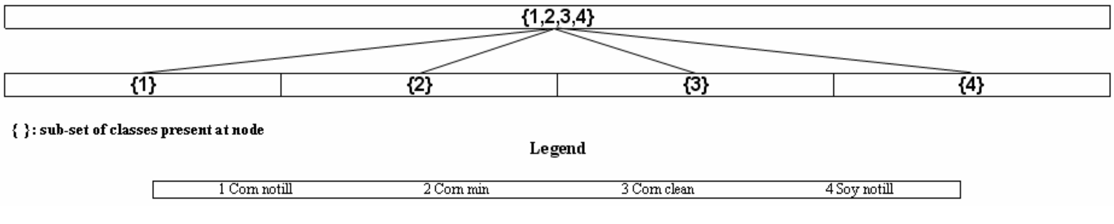

Figure 1 – Structure of single stage classifier

In the context of a single stage classifier, a problem that frequently arises deals with the selection of the variables to be used by the classifier. In order to prevent that the number of parameters to be estimated becomes too large, it is advisable to use a sub-set of variables, including the ones that show a higher discriminant power among the classes involved only. In the single stage approach all classes are considered simultaneously a fact that turns this selection into a difficult problem.

An another possible approach to the multiple class cases consists in the use of multiple stage classifiers, or decision tree classifiers, where the global problem is subdivided in smallest local problems. In this approach, the classification process of each pixel considers in each stage only a sub-set of classes. However, as the number of classes increases, so does the number of possible structures for the decision tree, turning the selection of the optimal tree structure into a difficult problem. A possible approach to solve this problem was proposed by Moraes (2005), which consists in adopting a pre-defined binary structure to the tree. In that work, the author proves the superiority of the defined structure over all others possible structures of binary tree classifiers.

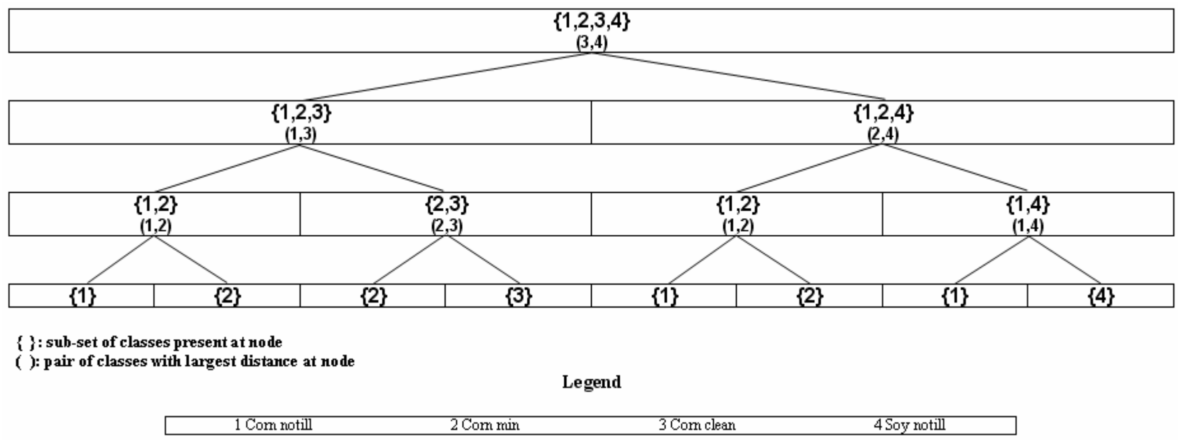

Figure 2 – Structure of multiple stage classifier

In the case of hierarchical discrimination between multiple classes, the multiple stage classifier structure is then defined as follows. First, at each stage of the classifier, statistical distances separating two classes at a time in node d are estimated by making use of the available training samples. In this study the Bhattacharyya distance (Moraes, 2005) was used. The pair of classes showing the largest distance is then selected to define the two descending nodes. The training samples available to the selected classes are entirely allocated into the corresponding descending nodes. The training samples belonging to the remaining classes in the parent node are classified into one of the descending nodes according with a decision rule. This process is then repeated until the terminal nodes – nodes including samples of a single class – are reached.

Once the structure of the binary multiple stage classifier is defined, as illustrated in the example shown in **Figure 2**. The actual classification of individual pixels in the image data can start. In each node pixels are compared to two classes only, namely, the pair previously selected by the largest statistical distance criterion, allowing the straight application of the formal logistic discrimination method. This process is applied sequentially across the tree, until the pixel reaches a terminal node in which case it is labeled according with this node.

4. Results

Hyper spectral image data acquired by the AVIRIS system with known ground truth was used to compare the efficiency of the single stage classifier against the accuracy provided by the multiple stage classifier approach. Out of the 224 bands available by Aviris sensor only 190 were considered in the experiments, due to the noisy effects of the atmospheric water vapor in the remaining bands. **Figure 3** presents the study area and the **Figure 4** shows the spectral signature of the four classes that were implemented into the classifier. A sample of 95 spectral bands, chosen systematically, was used to reduce the computational cost. The training sample size was chosen equal to 500 pixels per class. The same set of samples was also used for testing purposes (re-substitution method). All procedures of estimation were made in the software SPSS release 13.0

Figure 3 – Study area (composite color of AVIRIS image) and ground truth

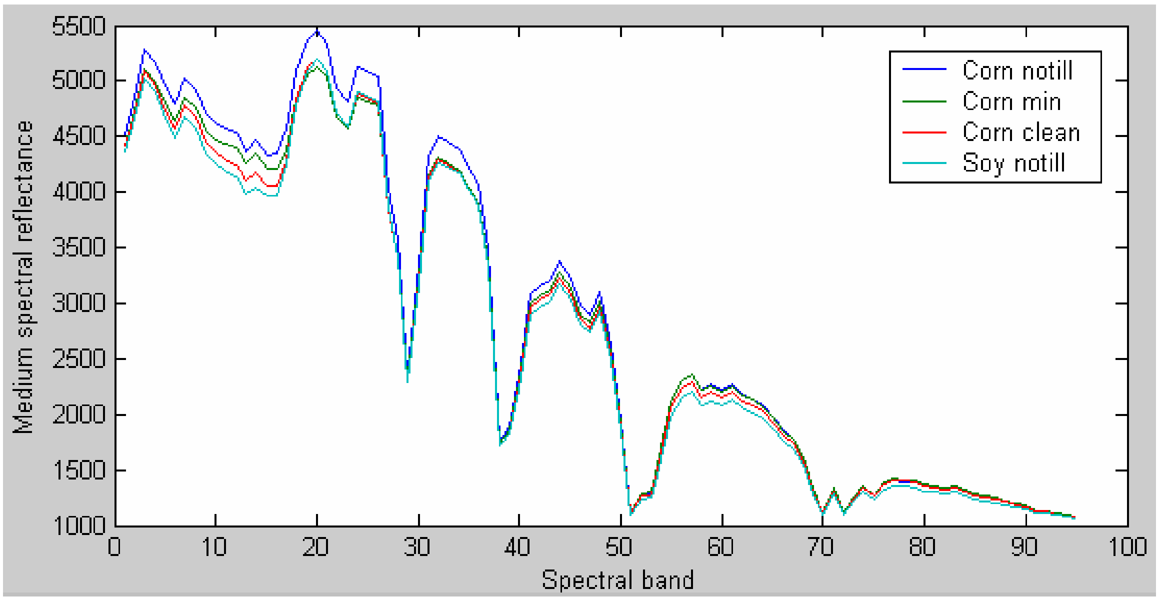

Figure 4 – Spectral signature to the four classes (corn notill; corn min; corn and soy notill)

In the experiments, the single stage classifier yielded 73,7% mean accuracy. As the classes in the experiments are spectrally very similar, there was confusion between them especially among the classes corn, corn min and corn notill. The soy notill class is the easiest to discriminate. The **Table 1** shows the results. The software SPSS needed about 10 minutes to realize this task in a PC AMD Athlon 3GHz with 512Mb.

Table 1 – Classification table using the logistic discrimination single stage classifier

	Predicted				
Observed	*Corn*	*Corn min*	*Corn notill*	*Soy notill*	*Percent Correct*
Corn	382	59	46	13	76,4%
Corn min	92	334	50	24	66,8%
Corn notill	25	62	362	51	72,4%
Soy notill	37	26	41	396	79,2%
Overall Percentage					*73,7%*

In the case of multiple stage classifier, there are in this experiments 96 parameters to be estimated at each node of decision tree, totalizing seven decision functions and 672 parameters. In the tree structured algorithm implemented in this experiment, there are eight different paths across the tree that a pixel can follow to reach a terminal node. The SPSS software doesn't make this procedure automatically, so a little program was written in SPSS language to solve this problem. There was necessary just few seconds to estimate the parameters at each stage, because the estimation process at the traditional binary logistic model is faster than multinomial logistic discrimination. The pixels' distribution among the nodes, using a 2000 pixels validation sample, is shown in **Figure 5.**

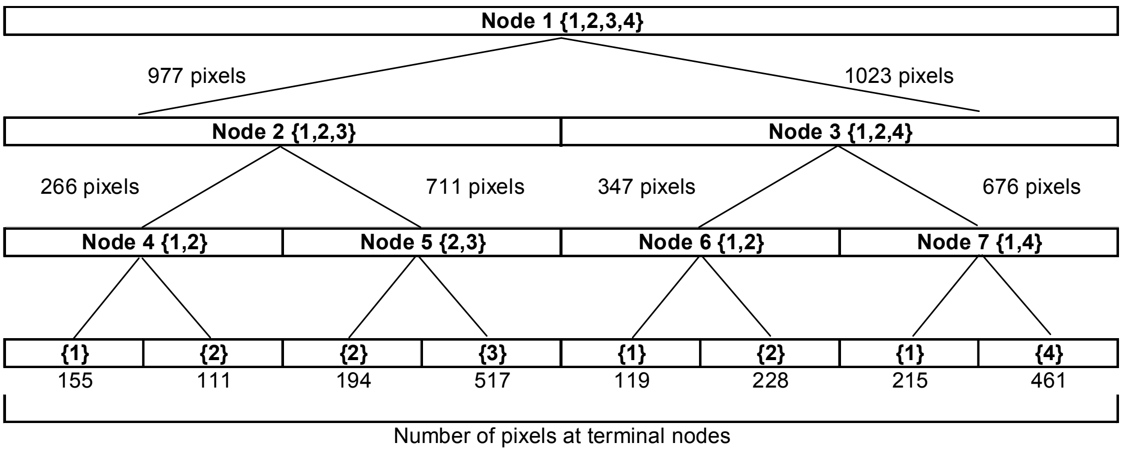

Figure 5 – The process of classification in multiple stage classifier

The mean accuracy yielded by the multiple stage classifier was slightly better in comparison of single stage. The final classification table obtained by using the validation sample is shown in **Table 2**.

Table 2 – Classification table using the logistic discrimination multiple stage classifier

	Predicted				
Observed	*Corn*	*Corn min*	*Corn notill*	*Soy notill*	*Percent Correct*
Corn	392	50	20	38	78,4%
Corn min	31	368	85	16	73,6%
Corn notill	24	78	382	16	76,4%
Soy notill	42	37	30	391	78,2%
Overall Percentage					*76,7%*

5. Final Remarks

The results found allow to make the following final remarks:

a) The multiple stage classifier presents better results than single one. The difference between them in the overall performance is not so great and the results agree with Moraes (2005) that found better results when a classification trees was used.

b) Regarding the processing time, the multiple stage classifier reached the estimates faster than single stage method.

c) Based on these results we suggest that the hierarchical classifiers can be used as an alternative to the single stage classifiers.

d) It is recommendable divide a classification complex problem in several simpler ones.

6. Acknowledgments

The authors acknowledge to *CNPq – Edital Universal* for its support.

References

Bittencourt, H. R. ; Clarke, R. T. Estudo comparativo entre o modelo de discriminação logístico e o método da máxima verossimilhança gaussiana. In: Simposio Latinoamericano de Percepción Remota, 9, 2000, Puerto Iguazú, Argentina. **Memorias**... Luján, SELPER, 2000. CD-ROM. Disponível em: <http://www.selper.org/trabajos/pro009.pdf>.

Bittencourt, H. R. ; Clarke, R. T. . Logistic Discrimination Between Classes with Nearly Equal Spectral Response in High Dimensionality. In: 2003 IEEE International Geoscience and Remote Sensing Symposium, 2003, Toulouse, France. **Annals…** 2003 IEEE IGARSS. Piscataway, NJ - USA: IEEE Operations Center, 2003a.

Bittencourt, H. R. ; Clarke, R. T. . Use of Classification and Regression Trees (CART) to Classify Remotely-Sensed Digital Images. In: 2003 IEEE International Geoscience and Remote Sensing Symposium, 2003, Toulouse, France. **Annals…** 2003 IEEE IGARSS. Piscataway, NJ - USA: IEEE Operations Center, 2003b.

Haertel, V.; Landgrebe, D. On the classification of classes with nearly equal spectral response in remote sensing hyperspectral image data**. IEEE Transactions on Geoscience and Remote Sensing**, v. 37, n. 5., pp. 2374-2386, 1999.

Jain, A. K.; Duin, R.P.; Mao J. Statistical Pattern Recognition: A Review. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, n. 1, pp. 04-37, 2000.

Moraes, D. A. O. **Extração de Feições em Dados Imagem com Alta Dimensão por Otimização da Distancia de Bhattacharyya em um Classificador de Decisão em Arvore**. Dissertação de Mestrado em Sensoriamento Remoto - PPGSR, Universidade Federal do Rio Grande do Sul, CNPq, 99 p., 2005.