Reprinted from

Symposium on

Machine Processing of

Remotely Sensed Data

June 3 - 5, 1975

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

IEEE Catalog No.
75CH1009-0 -C

RECURSIVE ESTIMATION OF PROPORTIONS IN EARTH OBSERVATIONS

Demetrios Kazakos

Rice University, Houston, Texas

ABSTRACT

In Earth observations problems, we usually have the situation of knowing accurately the probability density functions of the several classes of interest, and we need to classify a set of observations with unknown class proportions.

The observations are in n-dimensional space, where n is the number of spectral bands.

Two recursive algorithms for classifying the observations and estimating the prior probabilities are described. The first one achieves simultaneous classification of pixels and estimation of prior probabilities (or proportions) and the second one estimates the proportions in a recursive fashion.

There is a similarity of the second approach to maximum likelihood estimation, but the proposed method requires less computer time.

SUMMARY

FIRST METHOD

A. Two Class Case

We assume that each observation $X_i \in E^n$ comes from one of two hypotheses, H_1, H_2 with p.d.f's $f_1(X)$, $f_2(X)$.

We assume also that the prior probability π of H_1 is unknown.

Let P_n be the nth estimate of π.

Let

$$W_{k+1} = \begin{cases} 1 & \text{if } \dfrac{f_1(X_{k+1})}{f_2(X_{k+1})} \geq \dfrac{1 - P_k}{P_k} \\ 0 & \text{otherwise} \end{cases}$$

Extending the adaptive algorithm of Davisson and Schwartz [12] for simultaneous detection of the X_j observations and estimation of π, we construct the following recursive estimation-detection scheme:

$$P_{n+1} = P_n - (n+1)^{-1} \cdot \cdot [P_n - W_{n+1}]$$

i.e., at each step, we use the current estimate P_n for classification of X_{n+1} and determination of W_{n+1}.

It turns out that the above algorithm is simple but the estimate is biased. The contribution of the present work is the modification of the algorithm, so that

(a) The estimate P_n becomes unbiased.

(b) Convergence in the mean square sense is optimally accelerated.

This improvement is achieved by the introduction of two nonlinear transformations, $L(P)$ and $g(P)$, defined for $P \in (0,1)$.

The new algorithm is:

$$P_{n+1} = P_n - (n+1)^{-1} \cdot L(P_n) \cdot \cdot \left[g(P_n) - W_{n+1}\right]$$

Using Stochastic Approximation theory, the function g is chosen so that the estimate is unbiased. The choice that achieves this, is:

$$g(s) = \int_{R(s)} \left[s f_1(X) + (1-s) f_2(X) \right] dX$$

where

$$R(s) = \left\{ X \in E^n \; ; \; s f_1(X) \geq (1-s) f_2(X) \right\}$$

for $s \in (0,1)$

There is also an optimal choice for $L(P)$, in the sense that the asymptotic error variance is minimized.

The optimal choice is:

$$L(s) = \left[G(s)\right]^{-1}$$

for $s \in [e, 1-e]$

where

e = small pos. number

$$G(s) = \int_{R(s)} \left[f_1(X) - f_2(X) \right] dX$$

Using Sacks Stochastic Approximation theorem [8], convergence to the true parameter π is ensured.

The asymptotic error variance is:

$$\lim_{n \to \infty} n E(P_n - \pi)^2 = \left[G(\pi) \right]^{-2} \cdot g(\pi) \left(1 - g(\pi) \right)$$

B. $m > 2$ Class Case

Let $H_1, \ldots, H_m$ be the m hypotheses, with known probability density functions

$$f_1(X), \ldots, f_m(X) \ , \ X \in E^n$$

Let π = vector of prior probabilities

$$\pi = (\pi_1, \ldots, \pi_{m-1})$$

The extension of the original Davisson-Schwartz algorithm, is:

$$P_{n+1} = P_n - (n+1)^{-1} \cdot \left[P_n^1 - W_{n+1}^1, \ldots, P_n^{m-1} - W_{n+1}^{m-1} \right]$$

where

$$P_n = \left(P_n^1, \ldots, P_n^{m-1} \right)$$

and P_n^k is the nth estimate of π_k.

The quantities W_{n+1}^k are defined as follows:

$$W_{n+1}^k = \begin{cases} 1 \text{ if } P_n^k f_k(X_{n+1}) = \\ \quad = \max_j P_n^j f_j(X_{n+1}) \\ 0 \text{ otherwise} \end{cases}$$

They involve the "updating factors" in computing P_{n+1} from P_n.

Again, the algorithm is biased.

The improvement proposed is nontrivial extension of the $m = 2$ case.

We define the regions:

$$R_k(\pi) = \left\{ X \in E^n \ ; \ \pi_k f_k(X) = \max_s \pi_s f_s(X) \right\}$$

for $k = 1, \ldots, m$

Then, we define the functions:

$$g_k : (0,1)^m \to (0,1) \ ,$$

$k = 1, \ldots, m-1$

$$g_k(\pi) = \sum_{s=1}^{m} \pi_s \int_{R_s(\pi)} f_s(X) \, dX$$

for $k = 1, \ldots m-1$

Also, we define an $(m-1) \times (m-1)$ matrix,

$$L(\pi) = \left\{ L_{ij}(\pi) \right\} ,$$

$i, j = 1, \ldots, m-1$

$L_{ij}(\pi)$ are functions, defined on $(0,1)^m$

$$L_{ij}(\pi) : (0,1)^m \to (0, L_o)$$

Then, the proposed algorithm has the form:

$$P_{n+1} = P_n - (n+1)^{-1} \cdot A \cdot L(\pi) \cdot \left[g_1(P_n) - W_{n+1}^1, \ldots, g_{m-1}(P_n) - W_{n+1}^{m-1} \right]$$

The functions $g_k(\pi)$ make the algorithm unbiased.

Furthermore, there is a choice of the matrix $L(\pi)$ that will guarantee fast convergence, according to Stochastic Approximation theory of Sacks [8].

The choice of $L_{ij}(\pi)$ that will do it, is:

$$L_{sk}(\pi) = \int_{R_s(\pi)} \left[f_k(X) - f_m(X) \right] dX$$

$s, k = 1, \ldots, m-1$

A is a scalar constant, the adjustment of which will further improve convergence.

The formula for the asymptotic error convariance matrix

$$Q(\pi) = \lim_{n \to \infty} n\, E(P_n - \pi)^T \cdot (P_n - \pi)$$

is given in Ref. [4].

The algorithm is currently being applied to Earth observations data.

SECOND METHOD

Another way of estimating the prior probabilities is by the method of mixtures.

A. Two Calss Case

We construct the mixture density

$$g(X \mid \pi) = \pi f_1(X) + (1-\pi) f_2(X), \quad X \in E^n$$

The maximum likelihood estimate of π is the argument that maximizes

$$\sum_{k=1}^{N} \log g(X_k \mid \pi)$$

It is shown in [11] that the maximum likelihood estimate is asymptotically efficient for "nice" density functions.

However, it is in many cases difficult to compute because it involves the maximization of a Nth degree polynomial.

Instead, a recursive algorithm is proposed.

Let P_n be the nth estimate of π. Then the algorithm is:

$$P_{n+1} = P_n - (n+1)^{-1} \cdot L(P_n) \cdot \frac{f_1(X_{n+1}) - f_2(X_{n+1})}{P_n f_1(X_{n+1}) + (1-P_n) f_2(X_{n+1})}$$

where $L(P_n)$ is an adjustable weight function, defined on $P_n \in (0,1)$.

Using Sacks Stochastic Approximation theory, it is shown in [11] that P_n converges in mean square to π, (the true value).

Furthermore, there is a choice for $L(\pi)$ that will optimize the speed of convergence.

The optimal $L = L_o(\pi)$ is:

$$L_o(\pi) = \left[J(\pi) \right]^{-1}$$

where

$$J(\pi) = \int_{E^n} \left[f_1(X) - f_2(X) \right]^2 \cdot g^{-1}(X \mid \pi)\, dX$$

For the above choice, the estimate P_n is asymptotically efficient, i.e.

$$\lim n\, E(P_n - \pi)^2 = \left[J(\pi) \right]^{-1} =$$

Rao-Cramér lower bound.

B. m > 2 Class Case

We have to estimate the vector parameter $\pi = (\pi_1, \ldots, \pi_{m-1})$.

Let $P_n = (P_n^1, \ldots, P_n^{m-1}) =$ nth estimate of π.

Let

$$g(X \mid \pi) = \sum_{k=1}^{m-1} \pi_k f_k(X) + \left(1 - \sum_{k=1}^{m-1} \pi_k\right) f_m(X)$$

The algorithm now is:

$$P_{n+1} = P_n - (n+1)^{-1} \cdot g^{-1}(X_{n+1} \mid P_n) \cdot A(P_n) \cdot \left[f_1(X_{n+1}) - f_m(X_{n+1}), \ldots, f_{m-1}(X_{n+1}) - f_m(X_{n+1}) \right]$$

The details of the convergence properties of the above algorithm can be found in [11].

$A(P_n)$ is a function:

$$A(P) : (0,1)^{m-1} \to (0, A_o)$$

The adjustment of which accelerates convergence of the algorithm.

The above algorithm is similar in character to the maximum likelihood estimate of π.

It can be described as a recursive version of the MLE.

The updating of P_n is done by moving along the gradient of $\log g(X_{n+1} \mid \pi)$.

The proposed algorithm is currently being tested in comparison to the maximum likelihood one, using simulated and then ERTS data.

REFERENCES

[1] D. C. Boes, "On the Estimation of Mixing Distributions," Ann. Math. Statistics, 1966, p. 177.

[2] K. Fukunaga and T. Krile, "Calculation of Bayes' Recognition Error for Two Multivariate Gaussian Distributions," IEEE Trans. on Comp., March, 1969.

[3] T. Y. Young and G. Coraluppi, "Stochastic Estimation of a Mixture of Normal Density Functions Using Information Criterion," IEEE Trans. on IT, May, 1970.

[4] D. Kazakos, "Optimal Design of an Unsupervised Adaptive Classifier with Unknown Priors," ICSA, Rice University Technical Report, May, 1974. (Will appear as an article.)

[5] S. J. Yakowitz, "Unsupervised Learning and the Identifiability of Finite Mixtures," IEEE Trans. on IT, 1970, (3).

[6] H. Cramér, "Mathematical Methods of Statistics," Princeton University Press, 1946.

[7] M. D. Perlman, "The Limiting Behavior of Multiple Roots of the Likelihood Equation," Department of Statistics, University of Minnesota, Technical Report 125, July, 1969.

[8] J. Sacks, "Asymptotic Distribution of Stochastic Approximation Procedures," Ann. Math. Statistics, 1958 (2).

[9] J. R. Blum, "Multidimensional Stochastic Approximation Method," Ann Math. Statistics, 1954, p. 737.

[10] D. Sakrison, "Stochastic Approximation: A Recursive Method for Solving Regression Problems," Advances in Communication Systems, Vol. 2, A. V. Balakrishnan, ed. New York, Academic Press, 1966, pp. 51.

[11] D. Kazakos, "Recursive Estimation of Prior Probabilities Using Mixtures, "ICSA Rice Univ. Technical Report, Sept. 1974

[12] L. D. Davisson and S. C. Schwartz, "Analysis of a Decision Directed Receiver with Unknown Priors," IEEE Trans. on Inf. Th. 1970 (3).