

HIGHLY EFFECTIVE LOGISTIC REGRESSION MODEL FOR SIGNAL (ANOMALY) DETECTION

Dalton S. Rosario

Army Research Laboratory
Signal & Image Processing Division,
Adelphi, MD 20874

ABSTRACT

High signal to noise separation has been a long standing goal in the signal detection community. *High* in the sense of being able to separate orders of magnitude a signal(s) of interest from its surrounding noise, in order to yield a high signal detection probability at a near zero false-alarm rate. In this paper, I propose to use some of the advances made on the theory of logistic regression models to achieve just that. I discuss a logistic regression model—relatively unknown in our community—based on case-control data, also its maximum likelihood method and asymptotic behavior. An anomaly detector is designed based on the model's asymptotic behavior and its performance is compared to performances of alternative anomaly detectors commonly used with hyperspectral data. The comparison clearly shows the proposed detector's superiority over the others. The overall approach should be of interest to the entire signal processing community.

1. INTRODUCTION

Logistic regression models are commonly used in analyzing binary data which arise in studying relationships between disease and environment of genetic characteristics. After re-parameterization, logistic regression models can be expressed as a two-sample semiparametric model in which the log ratio of two density functions is linear in data, or in feature space.

Advances in this area produced adequacy examination of the hypothesized link for a given prospective sampling dataset [1], also a nonparametric regression method to test the validity of the logistic regression assumption [2], and some graphical methods for assessing logistic regression models [3]. More recently, the logistic regression assumption under a case-control sampling plan was tested [4] and its theory further formulated as an alternative approach to the one-way layout ANOVA (analysis of variance) that relaxes the normal assumption [5].

The focus in this paper is on using hyperspectral sensor imagery (HSI) with the mathematical statistic advances mentioned above to effectively achieve the automatic detection of certain types of signals (e.g., stationary ground vehicles) in natural clutter backgrounds (e.g., roads, grasses, trees).

Such an endeavor would benefit a military soldier, for instance, facing multiple tasks inside a remote ground control station by freeing this soldier from staring most of the time to a surveillance monitor. (Researchers at the Army Research Lab have extensive experience addressing these types of issues [6-8].)

The proposed approach has three fundamental parts: (1) introduction of a new method to obtain local sample variability from HIS; (2) transformation of samples into the discriminant metric SAM (spectral angle mapper) space, using the method in (1); and (3) introduction of a new data/feature model, relatively unknown in the signal processing community.

In this model, the probability distribution function (pdf) of transformed test samples is represented as a *distortion* of a reference pdf, which is also estimated from the incoming data. A new anomaly detector will be designed based on the model's asymptotic behavior and, for convenience, it shall be named *SemiP* detector.

The *SemiP* detector will assume no prior knowledge about the target and clutter statistic and no supervised training will be required. Its performance will be compared to performances of alternative anomaly detectors commonly used with hyperspectral data. The overall approach should be of interest to the entire signal processing community.

The paper is organized as follows: In Section 2, I describe a logistic regression model based on case-control data and its profiling maximum likelihood method. Section 3 describes an innovative method to obtain sample variability from the data and adaptation of theory to design the *SemiP* detector. Section 4 describes the hyperspectral dataset, other anomaly detectors, and comparison results. Section 5 summarizes the paper.

2. LOGISTIC REGRESSION MODEL

Let two vectors X_k have their components *iid* (independently, identically distributed) and let X_k be independent of X_j . Now, consider the model shown below, with $[.]^t$ denoting a transposed vector:

$$X_1 = [x_{11}, \dots, x_{1n_1}]^t \text{ iid } \sim g_1(x)$$

$$X_0 = [x_{01}, \dots, x_{0n_0}]^t \text{ iid } \sim g_0(x)$$

Regarding $g_l(x)$ as an exponential distortion of a reference $g_0(x)$, one can express

$$\frac{g_1(x)}{g_0(x)} = \exp(\alpha + \beta h(x)), \quad (1)$$

where $h(x)$ is an arbitrary but known function of x . (In this paper $h(x) = x$.) Letting a new vector t be constructed as

$$t = [x_{11}, \dots, x_{1n_1}, x_{01}, \dots, x_{0n_0}]^t \\ = [t_1, \dots, t_{n=n_1+n_0}]^t$$

and using (1), the MLE (maximum likelihood estimate) of α and β can be derived via the likelihood function

$$\zeta(\alpha, \beta, g_0) = \prod_{i=1}^{n=n_1+n_0} g_0(t_i) \prod_{j=1}^{n_1} \exp(\alpha + \beta h(x_{1j}))$$

Since g_0 is unknown, deriving MLE of α and β via standard procedures can not be achieved; however, a simple profiling method can be applied here to express g_0 in terms of α and β using the following facts: (i) fixing α and β , implies that maximizing ζ is equivalent to maximizing $\prod g_0$; (ii) g_0 has the properties of a pdf; and (iii) motivated by (1), an exponentially distorted g_0 is also a pdf. Using those facts, one can use any optimization method to maximize $\prod g_0$ (given a set of conditions from the three facts above), express g_0 in terms of α and β , and finally find the MLE via a set of score equations. A solution set should match the following one for $\rho = n_1/n_0$ [4]:

$$\frac{\partial \log[\zeta(\alpha, \beta, g_0)]}{\partial \alpha} = -\sum_{i=1}^n \frac{\rho \exp[\alpha + \beta h(t_i)]}{1 + \rho \exp[\alpha + \beta h(t_i)]} + n_1 = 0$$

$$\frac{\partial \log[\zeta(\alpha, \beta, g_0)]}{\partial \beta} = -\sum_{i=1}^n \frac{h(t_i) \rho \exp[\alpha + \beta h(t_i)]}{1 + \rho \exp[\alpha + \beta h(t_i)]} + \sum_{j=1}^{n_1} h(x_j) = 0.$$

3. DESIGNING THE SEMIP DETECTOR

To adapt the theory in Section 2 to the detection framework in context, a method to compare local information must be devised and the asymptotic behavior of the estimated parameters must be understood.

3.1. Obtaining Sample Variability

The most common method to compare local information in imagery is to use the traditional inside/outside window

implementation. The inside window represents the *test cell* and the outside window the *reference cell*. The notion is to check the likelihood of observed objects in the test cell of belonging to the class of observed surrounding background clutter in the reference cell. This approach, although very common, causes an unacceptably high number of false alarms on areas where nonhomogeneous clutter are observed in the reference cell. The problem occurs when samples from the reference cell are reduced to a set of statistics, hence, misrepresenting local regions consisting of mixed background clutter.

In this paper, I propose a more robust method to obtain local sample variability, which complements the strengths of the *SemiP* detector to suppress clutter that are highly nonhomogeneous. I define a third window (*variability cell*) and compute two sets of local features using this outer window: (1) discriminant metric between each sample from the variability cell and an average sample from the reference cell and (2) discriminant metric between these same variability-cell samples and the other average sample from the test cell. Feature set 1 will form the input reference to the new detector and feature set 2 will form the input test.

3.2. Theory Adapted for Detection

In reference to the theory in Section 2, since g_0 is a density, $\beta = 0$ implies $\alpha = 0$, and the hypothesis $H_0: \beta = 0$ infers that the populations of the two feature sets described in Section 3.1 are equally distributed, namely, $g_1 = g_0$. A realizable detector can now be designed from the following composite hypothesis test:

$$H_0: \beta = 0 \quad (g_1 = g_0) \quad \text{anomaly absent}$$

$$H_1: \beta \neq 0 \quad (g_1 \neq g_0) \quad \text{anomaly present}$$

Local regions in the entire imagery will be individually tested yielding a binary surface that depicts hypothesis H_1 as “1” and H_0 as “0.” An isolated object is expected to produce a cluster of “1” values (anomalies) in this surface.

The last two peaces to complete the *SemiP* detector is to find the pdf of $\hat{\beta}$ (MLE of β) and to select a discriminant metric (feature) for the hypothesis test. After consideration of many candidates, I settled for the metric SAM (spectral angle mapping) [9], which essentially computes the angle between two vectors. SAM is a popular metric in the hyperspectral community. In reference to the asymptotic behavior of $\hat{\beta}$, it can be shown [5] that

$$\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, \frac{\rho^{-1}(1+\rho)^2}{Var(t)}) \quad \text{as } n \rightarrow \infty$$

Using the above asymptotic behavior and the null hypothesis with $\beta_0 = 0$ as the true value of β , one can

construct a hypothesis test based on the chi-squared distribution with one degree of freedom (χ_1^2), or:

$$\chi = n\rho(1 + \rho)^{-2} \hat{\beta}^2 \hat{V}(t) \Rightarrow \chi_1^2$$

where the estimate of the test-statistic variance ($\hat{V}(t)$) is poorly approximated by the standard sample variance using the constructed vector t . A significant better approximation is to use the actual definition of variance (i.e., $E[(x-\mu)^2]$) and the resulting biased estimate of g_0 from the theory described in Section 2. With those comments, the *SemiP* detector is complete. High values of χ reject hypothesis H_0 , detecting anomalies.

4. EXPERIMENT RESULTS

4.1. Data

Experiment was carried out on data set from the hyperspectral digital imagery collection experiment (HYDICE) sensor. The data have a spatial resolution of approximately one meter. The HYDICE sensor records 210 spectral bands in the visible-to-near infrared (VNIR) and short-wave infrared (SWIR). The results shown in this section for one data cube are representative for the data set. An illustrative data cube is shown in Fig 1. The scene consists of six stationary motor vehicles (targets) in the presence of natural background clutter (e.g., trees, dirt road, grass). Due to window cell sizes, Fig 1 also shows the region in the scene tested for anomalies, which is depicted as a large rectangle, excluding one of the targets.

4.2. Other Detection Techniques

The *SemiP* detector will be compared to four other techniques: RX (reed-xiaoli), PCA (principle component analysis), EST (eigen separation transform), and FLD (fisher's linear discriminant). These techniques—or variations of them—are commonly used in the hyperspectral community. They are represented by the following equations [6]:

$$Score_{RX} = (\bar{x}_{in} - \bar{x}_{out})^t C_{out}^{-1} (\bar{x}_{in} - \bar{x}_{out})$$

$$Score_{PCA} = E_{in}^t (\bar{x}_{in} - \bar{x}_{out})$$

$$Score_{EST} = E_{\Delta C}^t (\bar{x}_{in} - \bar{x}_{out})$$

$$Score_{FLD} = E_{S_b/S_w}^t (\bar{x}_{in} - \bar{x}_{out})$$

where \bar{x}_{in} is a sample mean vector from a set of inside-window vectors; \bar{x}_{out} is similar but sampled from the outside window; C_{out}^{-1} is the inverse sample covariance

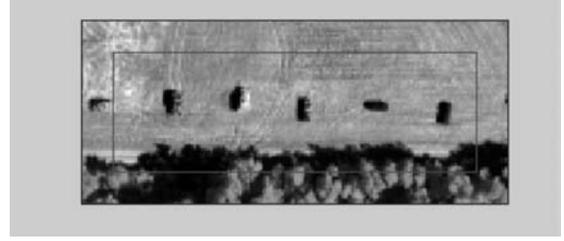


Figure 1 Nonhomogeneous, multicomponent scene from the HYDICE data collection. Typical anomaly detectors produce an unacceptable high number of false alarms (non-anomalies) in such a scene; local discontinuities degrade performance.

using all vectors sampled from the outside window; E_{in}^t is the highest energy eigenvector of the eigenvector decomposition of the inside-window covariance; $E_{\Delta C}^t$ is the highest positive energy eigenvector of the eigenvector decomposition of the covariance difference (inside-window minus outside-window); and E_{S_b/S_w}^t is the eigenvector decomposition of the scatter matrices ratio $S_B S_W^{-1}$, where

$$S_W = \sum_{i=1}^N (x_{in}^{(i)} - \bar{X}_{in})(x_{in}^{(i)} - \bar{X}_{in})^t + \sum_{i=1}^M (x_{out}^{(i)} - \bar{X}_{out})(x_{out}^{(i)} - \bar{X}_{out})^t$$

and

$$S_B = \sum_{i=1}^N (x_{in}^{(i)} - \bar{X}_{total})(x_{in}^{(i)} - \bar{X}_{total})^t + \sum_{i=1}^M (x_{out}^{(i)} - \bar{X}_{total})(x_{out}^{(i)} - \bar{X}_{total})^t$$

and \bar{X}_{total} is the sample average using samples from the inside and outside windows.

4.3. Results

ROC (receiver's operating characteristics) curves are used in this paper to quantify the differences in performance among the five techniques. The vertical axis of the ROC curves shown in Fig. 2 is the PD (probability of detection) domain and its horizontal axis is the FAR (false alarm ratio) domain in terms of the number of false alarms per a squared kilometer (km^2) area.

The quality of a detector can be readily assessed by noticing a key feature in the shape of its ROC curve. The closer the *knee* of a ROC curve is to the PD axis, the less sensitive the approach is to different decision thresholds, i.e., FAR does not change significantly as PD increases. As one can assess from Fig 2, the *SemiP* detector performs *dramatically* superior to the others.

This dramatic performance can be better appreciated by observing the decision surface (chi-squared values per location) in Fig 3. The surface was clipped at the value of 300, but it does continue to three orders of magnitude above this value. The five dominant peaks in that surface are the results produced by the five targets in the tested scene. Areas containing the presence of clutter mixtures

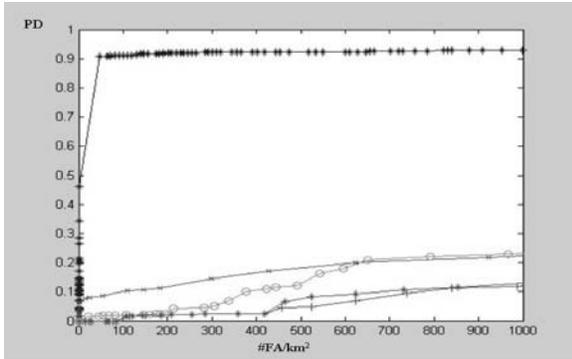


Figure 2 ROC curves of five anomaly detectors on scene depicted in fig. 1: *SemiP* (top), EST (2nd from top), RX (3rd from top), FLD (4th from top), and PCA (5th from top). The *SemiP* detector is significantly less sensitive to different decision thresholds.

(e.g., edge of terrain, edge of tree clusters), where other methods usually find a high number of false alarms (false anomalies) are significantly suppressed by the new approach. (Why?) Because, as part of the theory, features from different cells are not compared as two individual cells, they are combined and then compared to the test cell. This property in our hypothesis test produces more evidence that a subset of a clutter mixture (e.g., shadow) observed in the test cell might belong to the same distribution of the clutter mixture itself (e.g., trees and shadows), when observed in the reference cell. Performances of such cases are shown in Fig. 3 in the form of less-dominant peaks.

The set of combined feature samples is a spontaneous consequence of the mathematics as the logistic regression model is applied to our problem. Recall that in the model, the distribution functions of features in the SAM space are related by an exponential distortion.

5. FINAL REMARKS

A fully unsupervised anomaly detector (*SemiP*) has been presented for hyperspectral imagery. The approach has an adapted logistic regression model and a solution for its hypothesis test exploits recent advances in semiparametric theory. Performance of the *SemiP* detector in the visible to short-wave infrared region of the spectrum was compared to performances of four other techniques. The comparison clearly showed *SemiP* detector's superiority over the others. The asymptotic distribution of the test statistic under H_0 is independent of the unknown parameters, which implies that *SemiP* has the constant probability-of-error property. Having so, one can—in theory—select a decision threshold that yields a virtual zero probability of error. *Error* in this context means detection of non-anomalies, which is purely based on similarities between test samples and their immediate surroundings, not necessarily non-targets. This distinction should be noted.

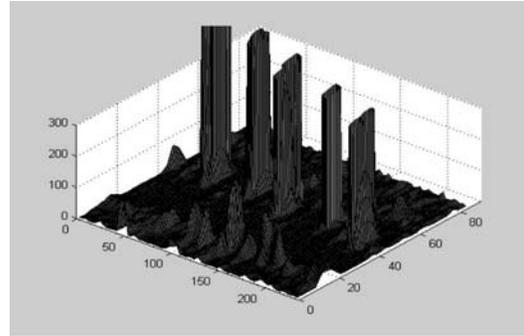


Figure 3 Decision surface produced by the *SemiP* detector for scene in fig. 1. The five dominant peaks represent undisputed anomalies in the tested scene. Notice other less-dominant peaks in the surface due to local discontinuities observed in fig. 1.

6. REFERENCES

- [1] D. Pregibon, "Goodness of link tests for generalized linear models," *Appl Statistics*, **29**, pp. 15-24, 1980.
- [2] A. Azzalini, A. Bowman, and W. Hardle, "On the use of nonparametric regression for model checking," *Biometrika*, **76**, pp. 1-11, 1989.
- [3] J.M. Landwehr, D. Pregibon, A.C. Shoemaker, "Graphical methods for assessing logistic regression models," *J. Am. Statist. Assoc.*, **79**, pp. 61-83, 1984.
- [4] J. Qin and B. Zhang, "A goodness of fit test for logistic regression models based on case-control data," *Biometrika*, **84**, pp. 609-618, 1997.
- [5] K. Fokianos, B. Kedem, and J. Qin, "A semiparametric approach to the one-way layout," *Technometrics*, pp. 56-65, 2001.
- [6] H. Kwon, S.Z. Der, and N.M. Nasrabadi, "Adaptive anomaly detection using subspace separation for hyperspectral imagery," *Opt. Eng.*, **42** (11), November 2003.
- [7] D.S. Rosario, "Simultaneous Multi-Target Detection and False Alarm Mitigation Algorithm for the Predator UAV TESAR ATR System." *Procs. 10th Int'l. Conf. Image Analysis & Proc.*, IEEE Computer Soc, Venice, Italy, pp. 577-581, Sept 1999.
- [8] D.S. Rosario, "Estimating Squinted SAR Data: An Efficient Multivariate Minimization Approach Using Essential 3-D Target Information." *Proc. IEEE Int'l. Conf. Image Proc.*, Santa Barbara, CA, October 1997.
- [9] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Magazine*, pp 29., Jan. 2002.