

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

Copyright © 1975 IEEE
The Institute of Electrical and Electronics Engineers, Inc.

Copyright © 2004 IEEE. This material is provided with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the products or services of the Purdue Research Foundation/University. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org.

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

SINGLE-CLASS CLASSIFICATION*

T. C. Minter

Lockheed Electronics Company, Inc.
Aerospace Systems Division
Houston, Texas

I. ABSTRACT

Often, when classifying multispectral data, only one class or crop is of interest, such as wheat in the Large Area Crop Inventory Experiment (LACIE). Usual procedures for designing a Bayes classifier require that labeled training samples and therefore ground truth be available for the "class of interest" plus all confusion classes defined by the multispectral data. This paper will consider the problem of designing a two-class Bayes classifier which will classify data into the "class of interest" or the "other" classes but will require only labeled training samples from the "class of interest" to design the classifier. Thus, this classifier minimizes the need for ground truth. For these reasons, the classifier is referred to as a single-class classifier. A procedure for evaluating the overall performance of the single-class classifier in terms of the probability of error will be discussed.

II. INTRODUCTION

Often, when classifying multispectral data, only one class or crop is of interest such as wheat in the LACIE. Usual procedures for designing a Bayes classifier require that labeled training samples and therefore ground truth be obtained for the "class of interest" and all confusion classes in the multispectral data. Class statistics are estimated from these labeled samples. In the discussions that follow, a Bayes classifier will be presented which classifies samples into the "class of interest" or the "other" classes but requires only labeled training samples for the "class of interest" to design the classifier. Thus, this classifier minimizes the need for ground truth. For these reasons, the classifier will be referred to as a single-class classifier.

In section IV the problem of evaluating the performance of the single-class classifier in terms of the overall probability of misclassification will be discussed.

III. THE SINGLE-CLASS CLASSIFIER

It is well known that the Bayes decision rule for minimizing the probability of misclassification in the two-class case is

Decide X is a member of class 1 if

$$q_1 p(X/1) \geq q_0 p(X/0) \quad (1)$$

Otherwise X is a member of class 0

where q_1 and q_0 are the class *a priori* probabilities and $p(X/1)$ and $p(X/0)$ are the

*This work was funded by NASA under contract NAS 9-12200.

class conditional density functions. In this case the "class of interest" corresponds to class 1, and class \emptyset corresponds to the "other" class. In the discussions that follow, it is assumed that estimates are available for q_1 and $p(X/1)$. In addition, $p(X/1)$ is assumed to be normal, and class \emptyset is assumed to be a mixture of normal density functions

$$q_{\emptyset}p(X/\emptyset) = \sum_{i=2}^m q_i p(X/i) \quad (2)$$

where $\{q_2, q_3, \dots, q_m\}$ are the *a priori* probabilities of each of the subclasses of "other." In addition

$$q_{\emptyset} = q_2 + q_3 + \dots + q_m \quad (3)$$

and $\{p(X/2), p(X/3), \dots, p(X/m)\}$ are the class density functions for the subclasses of "other." In the above expressions X is a random, independent, n -dimensional measurement vector; i.e., $X = \{X_1, X_2, \dots, X_n\}$. The overall mixture density is given by

$$p(X) = \sum_{i=1}^m q_i p(X/i) \quad (4)$$

or

$$p(X) = q_1 p(X/1) + \sum_{i=2}^m q_i p(X/i) \quad (5)$$

Substituting equation 2 into equation 5,

$$p(X) = q_1 p(X/1) + q_{\emptyset} p(X/\emptyset) \quad (6)$$

Therefore, from equation 6, $q_{\emptyset} p(X/\emptyset)$ can be defined as

$$q_{\emptyset} p(X/\emptyset) = p(X) - q_1 p(X/1) \quad (7)$$

By substituting equation 7 into equation 1, the Bayes decision rule for classifying a measurement vector X as class 1 (i.e., the class of interest) is

$$q_1 p(X/1) \geq p(X) - q_1 p(X/1) \quad (8)$$

or rearranging equation 8,

Decide X is a member of class 1
(i.e., the class of interest) if

$$q_1 p(X/1) \geq 1/2 p(X) \quad (9)$$

otherwise, X is a member of class \emptyset

The density $p(X)$ can theoretically be estimated without the use of any training samples. Thus, training samples are needed only to estimate $p(X/1)$. The *a priori* probability q_1 possibly can be estimated from historical information. This procedure was developed by J. A. Quirein (ref.).

IV. ESTIMATING THE PROBABILITY MISCLASSIFICATION FOR THE SINGLE-CLASS CLASSIFIER

The performance of a classifier as measured in terms of the probability of misclassification is an important indicator of the quality of the classification results. Assuming labeled samples are available only for the "class of interest" (i.e., class 1),

then only the probability of misclassifying samples from class 1 into class \emptyset [i.e., $\Pr(\emptyset/1)$] may be estimated directly.

The lack of labeled samples for the "other" class causes difficulty in evaluating the probability of misclassifying samples from class \emptyset into class 1 by usual procedures. In the discussions that follow a procedure will be presented for estimating the overall probability of error for the single-class classifier using only the labeled samples from the "class of interest."

By definition the Bayes decision rule in equation 1 defines a Bayes region R_1 for class 1 and a set of vectors X^1 of N_1 samples; i.e.,

$$R_1 = \{X | q_1 p(X/1) \geq q_\emptyset p(X/\emptyset)\} = \{(X_1^{(1)}, X_2^{(1)}, X_3^{(1)}, \dots, X_{N_1}^{(1)})\} \quad (10)$$

Similarly, the Bayes region for class \emptyset is R_\emptyset , which is defined as

$$R_\emptyset = \{X | q_\emptyset p(X/\emptyset) \geq q_1 p(X/1)\} \quad (11)$$

The probability of error P_e is by definition

$$P_e = \int_{R_\emptyset} q_1 p(X/1) dx + \int_{R_1} q_\emptyset p(X/\emptyset) dx \quad (12)$$

Substituting equation 7 into equation 12,

$$P_e = \int_{R_\emptyset} q_1 p(X/1) dx + \int_{R_1} [p(X) - q_1 p(X/1)] dx \quad (13)$$

or

$$P_e = q_1 \int_{R_\emptyset} p(X/1) dx - q_1 \int_{R_1} p(X/1) dx + \int_{R_1} p(X) dx \quad (14)$$

Since the conditional probability of misclassifying samples from class 1 into class \emptyset is

$$\Pr(\emptyset/1) = \int_{R_\emptyset} p(X/1) dx \quad (15)$$

and the conditional probability of correct classification for class 1, $\Pr(1/1)$, is

$$\Pr(1/1) = \int_{R_1} p(X/1) dx \quad (16)$$

and the probability of a sample being an element of region R_1 , $\Pr(X \in R_1)$, is

$$\Pr(X \in R_1) = \int_{R_1} p(X) dx \quad (17)$$

then by substituting equations 15, 16, and 17 into equation 14, the probability of error can be defined as

$$P_e = q_1 [\Pr(\emptyset/1) - \Pr(1/1)] + \Pr(X \in R_1) \quad (18)$$

Thus, from equation 18 above, the probability of misclassification may be estimated using only the labeled samples from class 1, i.e., the "class of interest." A Monte Carlo technique may be used to estimate $\Pr(0/1)$ and $\Pr(1/1)$ [i.e., by empirically counting the number of training (or test) samples from class 1 that are correctly classified and misclassified]. $\Pr(X \in R_1)$ can be estimated from unlabeled samples; i.e.,

$$E\{\Pr(X \in R_1)\} = \frac{N_1}{N_T} \quad (19)$$

where N_T is the total number of samples classified, and N_1 is the number of samples classified as members of class 1 (see equation 10). The *a priori* probability of class 1 must be known or perhaps estimated from historical data.

V. CONCLUSIONS

A single-class maximum likelihood classifier has been presented in this paper. The classifier classifies data into the "class of interest" or the "other" class. The classifier requires only labeled training samples and therefore ground truth for the "class of interest." This procedure minimizes the need for ground truth. In addition, a procedure for estimating the overall performance of the single-class classifier using only labeled samples from the "class of interest" has been presented.

VI. REFERENCES

- J. A. Quirein, 1974: "Estimation of the Density Function for Non-Wheat," LEC/ASD Internal Memorandum, ref. 642-1254.