DEPARTMENT OF MATHEMATICS
UNIVERSITY OF HOUSTON HOUSTON, TEXAS

CR-151130

(NASA-CR-151130) FEATURE COMBINATIONS AND THE BHATTACHARYYA CRITERION (Houston Univ.) 11 p HC A02/MF A01 CSCL 12A

N77-13759

Unclas
G3/65 56937

FEATURE COMBINATIONS AND THE BHATTACHARYYA CRITERION

BY H.P. DECELL AND S.K. MARANI
REPORT #55 JUNE 1976

PREPARED FOR
EARTH OBSERVATION DIVISION, JSC
UNDER
CONTRACT NAS-9-15000

HOUSTON, TEXAS 77004

FEATURE COMBINATIONS AND THE BHATTACHARYYA CRITERION

by

Henry P. Decell, Jr. and Salma K. Marani
Department of Mathematics
University of Houston
Houston, Texas

Report #55
NASA Contract NAS-9-15000

FEATURE COMBINATIONS AND THE BHATTACHARYYA CRITERION

Henry P. Decell, Jr. and Salma K. Marani

Department of Mathematics
University of Houston

ABSTRACT

We develop a procedure for calculating a kxn rank k matrix B for data compression using the Bhattacharyya bound on the probability of error and an iterative construction using Householder transformations. Two sets of remotely sensed agricultural data are used to demonstrate the application of the procedure. The results of the applications give some indication of the extent to which the Bhattacharyya bound on the probability of error is affected by such transformations for multivariate normal populations.

1. INTRODUCTION

For n-dimensional normal classes $N(\mu_i \Sigma_i)$ $i = 1,\ldots,m$, the Bhattacharyya coefficient (Andrews, 1972) for class i and j is given by:

$$\rho(i,j) = (q_i q_j)^{\frac{1}{2}} \int_{R^n} \{p_i(x) p_j(x)\}^{\frac{1}{2}} dx$$

and the Bayes probability of error (Anderson, 1958) (Andrews, 1972) by

$$P_e = 1 - \int_{R^n} \max_{1 \le i \le m} \{q_i p_i(x)\} dx$$

where $p_i(x)$ denotes the conditional density of the random variable X given that $X \sim N(\mu_i, \Sigma_i)$ and $q_1, \ldots, q_m$, respectively, denote the (known) a priori probabilities of the classes $N(\mu_i \Sigma_i)$ $i = 1, \ldots, m$.

It has been shown (Andrews, 1972) (Kaileth, 1967) that

$$P_e \le \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \{q_i q_j\}^{\frac{1}{2}} \int_{R^n} \{p_i(x) p_j(x)\}^{\frac{1}{2}} dx$$

If one considers a kxn rank k linear transformation B of the random variable X (i.e., $Y \equiv BX$), then the Bhattacharyya coefficient for class i and j for the classes $N(B\mu_i, B\Sigma_i B^T)$, $i = 1, \ldots, m$ is:

$$\rho_B(i,j) \equiv \{q_i q_j\}^{\frac{1}{2}} \int_{R^k} \{p_i(y,B) p_j(y,B)\}^{\frac{1}{2}} dy$$

and the Bayes probability of error for the classes $N(B\mu_i, B\Sigma_i B^T)$, $i = 1, \ldots, m$ is:

$$P_e(B) = 1 - \int_{R^k} \max_{1 \le i \le m} \{p_i(y,B)\} dy$$

where $p_i(y,B)$, $i = 1, \ldots, m$ denotes the conditional density of the random variable $Y = BX$ given that $Y \sim N(B\mu_i, B\Sigma_i B^T)$. It follows, since $P_e \leq \rho \equiv \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \rho(i,j)$, that

$$P_e(B) \leq \rho(B) \equiv \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \rho_B(i,j)$$

and moreover, (Decell and Quirein, 1973) (Kaileth, 1967), that

(1) $P_e \leq P_e(B) \leq \rho(B)$.

(2) $P_e = P_e(B)$ if and only if $\rho = \rho(B)$.

2. THEORETICAL PRELIMINARIES

Let k be an integer $(0 < k < n)$, and $N(\mu_i, \Sigma_i)$ $i = 1,\ldots,m$ be n-variate normal populations with a priori probabilities $q_1,\ldots,q_m$. We would like to construct a kxn rank k matrix B that will minimize $\rho(B)$. The theoretical extent to which this is possible and the basis for the construction (Decell and Smiley, to appear) is summarized in the following theorem. Let $C = \{u \in R^n : ||u|| = 1\}$ and $T(H) = \{H = I - 2uu^T : u \in C\}$ denote the set of Householder transformations on R^n (Householder, 1958).

Theorem. For each positive i, let $H_i \in T(H)$ be chosen such that

$$\rho((I_k|Z)H_1) = \underset{H \in T(H)}{\text{g.l.b}}\ \rho((I_k|Z)H)$$

and

$$\rho((I_k|Z)H_{i+1}H_i \cdots H_1) = \underset{H \in T(H)}{\text{g.l.b.}}\ \rho((I_k|Z)HH_i \cdots H_1)$$

then,

(1) $\rho((I_k|Z)H_{i+1}H_i \cdots H_1) \leq \rho((I_k|Z)H_i \cdots H_1)$.

(2) $\rho((I_k|Z)H_{i+1} \cdots H_1) \leq \rho((I_k|Z)H_i \cdots H_1 H,\ H \in T(H))$.

(3) $\rho((I_k|Z)H_{i+1}H_i \cdots H_1) \leq \rho((I_k|Z)HH_i \cdots H_1,\ H \in T(H))$.

(4) $\rho((I_k|Z)H \cdots H_{i-(p-1)}HH_{i-(p+1)}H_1) \leq \rho((I_k|Z)H_{i+1}H_i \cdots H_1),\ H \in T(H)$ and $p = 0,\ldots,i-2$.

(5) The monotone sequence of real numbers $\{\rho(B_i)\}_{i=1}^{\infty}$ where

$B_i = (I_k|Z)H_i \cdots H_1$ is bounded below by P_e and hence

$$\lim_{i\to\infty} \rho(B_i) = \underset{i}{\text{g.l.b.}}\{\rho(B_i)\}$$

We know (Decell and Quirein, 1973) that there is some kxn rank k matrix, say $\hat{B}$, that minimizes $\rho(B)$. If $\rho(B) < \underset{i}{\text{g.l.b.}}\{\rho(B_i)\}$ we will call the sequence $\{B_i\}_{i=1}^{\infty}$ <u>sub optimal</u> (<u>optimal</u> in the case of equality). There are several results (Decell and Smiley, to appear) that lend credibility to the conjecture that the sequence is optimal and cofinally constant beyond the index $i = \min\{k,n-k\}$. We will proceed with the development of an iterative procedure for constructing the subject sequence and, finally, tabulate results of applications to remotely sensed agricultural data with equal <u>a priori</u> class probabilities. The approach (and its merit) will depend upon the bound provided by the inequality $P_e \leq \rho(B_i)$ $i = 1,2,\ldots$, the non-increasing nature of the sequence $\{\rho(B_i)\}_{i=1}^{\infty}$, and the ability to manipulate the expressions for $\rho(B_i)$, $i = 1,2,\ldots$ in the case of normal populations.

3. THE GRADIENT OF $\rho((I_k|Z)H)$

We will develop an expression (for the case of normal n-variate populations $N(\mu_i,\Sigma_i)$, $i = 1,\ldots,m$) for the gradient of $\rho((I_k|Z)H)$ where $H \in T(H)$ has the form $H = I-2\frac{xx^T}{x^Tx}$, $x \neq \theta$.

This expression will be used in a steepest descent procedure to calculate each Householder transformation H_1, H_2, $H_3,\ldots$ described in the preceding theorem. For m populations $N(\mu_i\Sigma_i)$, $i = 1,\ldots,m$ it is easy to establish that in order to calculate H_{i+1}, one need only apply the steepest descent procedure to the Bhattacharyya coefficient determined by the populations $N(H_i \cdots H_1\mu_j, H_i \cdots H_1\Sigma_j H_1 \cdots H_i)$ $j = 1,\ldots,m$.

The expression for $\rho_{(I_k|Z)H}(i,j)$ is given by (Andrews, 1972) (Kaileth, 1967) (for the case of equal a priori probabilities $q_i = 1/m$, $i = 1,\ldots,m$):

$$\rho_{(I_k|Z)H}(i,j) = \frac{1}{m}\exp - \frac{1}{4}\delta_{ij}^T(\Sigma_i+\Sigma_j)^{-1}\delta_{ij} - \frac{1}{2}\ln\left(\frac{|\hat{\Sigma}_i+\hat{\Sigma}_j|}{2^k|\hat{\Sigma}_i|^{1/2}|\hat{\Sigma}_j|^{1/2}}\right)$$

where $\hat{\delta}_{ij} = (I_k|Z)H(\mu_i-\mu_j)$ and $\hat{\Sigma}_i = (I_k|Z)H\Sigma_i H(I_k|Z)^T$, in which case,

$$\rho((I_k|Z)H) = \sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\rho_{(I_k|Z)H}(i,j).$$

If we define

$$F_{ij} = -\frac{1}{4}\hat{\delta}_{ij}^T(\hat{\Sigma}_i+\hat{\Sigma}_j)^{-1}\hat{\delta}_{ij} \quad \text{and} \quad G_{ij} = -\frac{1}{2}\ln\left(\frac{|\hat{\Sigma}_i+\hat{\Sigma}_j|}{2^k|\hat{\Sigma}_i|^{1/2}|\hat{\Sigma}_j|^{1/2}}\right)$$

we have that the differential of $\rho_{(I_k|Z)H}(i,j)$ is

$$d(\rho_{(I_k|Z)H}(i,j)) = \frac{1}{m}\exp(F_{ij}+G_{ij})(d(F_{ij}) + d(G_{ij})).$$

from whence it follows that

$$d(\rho((I_k|Z)H)) = \frac{1}{m}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\exp(F_{ij}+G_{ij})(d(F_{ij}) + d(G_{ij})).$$

In order to simplify the notation, define $\Sigma_{ij} = \Sigma_i + \Sigma_j$ and $\Delta_{ij} = (\mu_i-\mu_j)(\mu_i-\mu_j)^T$.

Let $tr(\cdot)$ denote the trace of $(\cdot)$ and $|\cdot| = \det(\cdot)$. With a bit of matrix algebra it follows that

$$F_{ij} = -\frac{1}{4}tr\{((I_k|Z)H\Sigma_{ij}H(I_k|Z)^T)^{-1}(I_k|Z)H\Delta_{ij}H(I_k|Z)^T\}$$

and

$$G_{ij} = -\frac{1}{2}\ln|(I_k|Z)H\Sigma_{ij}H(I_k|Z)^T| + \frac{1}{4}\ln|(I_k|Z)H\Sigma_i H(I_k|Z)^T|$$
$$+\frac{1}{4}\ln|(I_k|Z)H\Sigma_j H(I_k|Z)^T| + \frac{k}{2}\ln 2.$$

We will now develop expressions for $d(F_{ij})$ and $d(G_{ij})$, $i,j = 1,\ldots,m$.

According to Decell and Quirein (1973)

$$d(F_{ij}) = -\frac{1}{2}\operatorname{tr}\{d((I_k|Z)H)Q_{ij}\}$$

where $B = (I_k|Z)H$ and

$$Q_{ij} = [\Delta_{ij}B^T - \Sigma_{ij}B^T(B\Sigma_{ij}B^T)^{-1}B\nabla_{ij}B^T](B\Sigma_{ij}B^T)^{-1}.$$

Since $H = I - 2\frac{xx^T}{x^Tx}$ it follows that

$$d((I_k|Z)H) = d((I_k|Z)(I - 2\frac{xx^T}{x^Tx})) = -2(I_k|Z)d\left(\frac{xx^T}{x^Tx}\right)$$

$$= -2(I_k|Z)\left\{\frac{x^Txd(xx^T) - xx^Td(x^Tx)}{(x^Tx)^2}\right\}$$

$$= \frac{-2(I_k|Z)}{(x^Tx)^2}\{x^Tx(d(x)x^T+xd(x)^T)-xx^T(d(x)^Tx+x^Td(x))\}$$

$$= \frac{-2(I_k|Z)}{(x^Tx)^2}\{(d(x)x^Txx^T+xx^Txd(x)^T-xx^Td(x)x^T-xd(x)^Txx^T\}$$

$$= \frac{-2(I_k|Z)}{(x^Tx)^2}\{(d(x)x^T-xd(x)^T)xx^T-xx^T(d(x)x^T-xd(x)^T)\}.$$

Substituting the latter in the expression

$$d(F_{ij}) = -\frac{1}{2}\,tr\{d((I_k|Z)H)Q_{ij}\}$$

and using the fact that $tr(AB) = tr(BA)$, we have

$$d(F_{ij}) = -\frac{1}{2}tr\left\{\frac{-2(I_k|Z)}{(x^Tx)^2}\,[(d(x)x^T - xd(x)^T)xx^T - xx^T(d(x)x^T - xd(x)^T)]\,Q_{ij}\right\}$$

$$= \frac{1}{(x^Tx)^2}tr\{Q_{ij}(I_k|Z)[(d(x)x^T - xd(x)^T)xx^T - xx^T(d(x)x^T - xd(x)^T)]\}$$

$$= \frac{1}{(x^Tx)^2}\,tr\{xx^TQ_{ij}(I_k|Z)(d(x)x^T - xd(x)^T) - Q_{ij}(I_k|Z)xx^T(d(x)x^T$$

$$-xd(x)^T)\}.$$

With a little matrix algebra (and some patience) it follows that

$$d(F_{ij}) = \frac{1}{(x^Tx)^2}\,tr\{[(xx^TQ_{ij}(I_k|Z) - Q_{ij}(I_k|Z)xx^T)^T$$

$$-(xx^TQ_{ij}(I_k|Z) - Q_{ij}(I_k|Z)xx^T)]xd(x)^T\}$$

We now find an expression for $d(G_{ij})$. First, recall (Kullback, 1968) that

$$d(\ln|B\Sigma B^T|) = 2tr\{d(B)\Sigma B^T(B\Sigma B^T)^{-1}\}$$

so that

$$d(G_{ij}) = -tr\{d((I_k|Z)H)\Sigma_{ij}H(I_k|Z)^T((I_k|Z)H\Sigma_{ij}H(I_k|Z)^T)^{-1}\}$$

$$-\frac{1}{2}tr\{d((I_k|Z)H)\Sigma_iH(I_k|Z)^T((I_k|Z)H\Sigma_iH(I_k|Z)^T)^{-1}$$

$$+\frac{1}{2}\,tr\{d((I_k|Z)H)\Sigma_jH(I_k|Z)^T((I_k|Z)H\Sigma_jH(I_k|Z)^T)^{-1}\}.$$

Obviously, the summands in the expression for $d(G_{ij})$ differ from the expression

$$d(F_{ij}) = -\frac{1}{2}\, tr\{d((I_k|Z)H)\, Q_{ij}\}$$

only by multiplicative constants and the matrix Q_{ij}. Hence, we may use the final expression for $d(F_{ij})$ to obtain the expression for $d(G_{ij})$ by simply adjusting the multiplicative constants and replacing Q_{ij} (in each summand in $d(G_{ij})$) with the expressions

$$J_{ij} = \Sigma_{ij}H(I_k|Z)^T[(I_k|Z)H\Sigma_{ij}H(I_k|Z)^T]^{-1}$$

$$K_{ij} = \Sigma_i H(I_k|Z)^T[(I_k|Z)H\Sigma_i H(I_k|Z)^T]^{-1}$$

$$L_{ij} = \Sigma_j H(I_k|Z)^T[(I_k|Z)H\Sigma_j H(I_k|Z)^T]^{-1}$$

At this point we will simplify the notation. Let

$$\hat{Q}_{ij} = (xx^TQ_{ij}(I_k|Z)-Q_{ij}(I_k|Z)xx^T)^T-(xx^TQ_{ij}(I_k|Z)-Q_{ij}(I_k|Z)xx^T)$$

and let $\hat{J}_{ij}$, $\hat{K}_{ij}$, and $\hat{L}_{ij}$ be similarly defined by substituting, respectively, J_{ij}, K_{ij}, and L_{ij} for Q_{ij} in the expression for $\hat{Q}_{ij}$, $i,j = 1,\ldots,m$. It follows that

$$d(F_{ij}) = \frac{1}{(x^Tx)^2}\, tr(\hat{Q}_{ij}xd(x)^T)$$

$$d(Gij) = \frac{2}{(x^Tx)^2}\, tr(\hat{J}_{ij}xd(x)^T) - \frac{1}{(x^Tx)^2}\, tr(\hat{K}_{ij}xd(x)^T) - \frac{1}{(x^Tx)^2}\, tr(\hat{L}_{ij}xd(x)^T).$$

In order that x be extremal, it is sufficient that x satisfy

$$G(x) \equiv \frac{1}{m} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \frac{\exp(F_{ij}+G_{ij})}{(x^T x)^2} (\hat{Q}_{ij} + 2\hat{J}_{ij} - \hat{K}_{ij} - \hat{L}_{ij})x = \theta.$$

Of course, the function $G(x)$ is the gradient of

$\rho((I_k|Z)(I - 2\frac{xx^T}{x^T x}))$ with respect to x.

With $G(x)$, we use a steepest descent technique to construct H_1. The process is repeated for the construction of H_2 since, given H_1, the problem of constructing H_2 is identical to that of constructing H_1 provided the populations are taken to be $N(H_1\mu_i, H_1\Sigma_i H_1)$ $i = 1,\ldots,m$.

Test results are presented in the following tables for nine twelve channel, C-1 flight line agricultural classes: soybeans, corn, oats, red-clover, alfalfa, rye, bare soil, and two types of wheat. The Hill County data is sixteen channel data for five agricultural classes: winter wheat, fallow crop, barley, grass, and stubble.

C-1 FLIGHT LINE DATA

$n = 12$, $m = 9$, $k = 6$, $\rho = .024$

Iteration	H_{B_1}	H_{B_2}	H_{B_3}
0	.327	.109	.134
1	.223	.060	.034
2	.171	.062	.033
3	.135	.068	.032
4	.116	.058	.031
5	.1157	.055	.0309
6	.1150	.054	.0303

HILL COUNTY DATA

n = 16, m = 5, k = 6. ρ = .107

Iteration	H_{B_1}	H_{B_2}	H_{B_3}
0	.872	.336	.299
1	.785	.310	.287
2	.525	.286	.232
3	.439	.273	.227
4	.576	.267	.226
5	.386	.265	.224
6	.363	.264	.223

BIBLIOGRAPHY

Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons, Inc.

Andrews, H.C. (1972). Introduction to Mathematical Techniques in Pattern Recognition. New York: Wiley-Interscience.

Decell, H.P. Jr. and Quirein, J.A. (March, 1973). "An Iterative Approach to the Feature Selection Problem". IEEE Cat. #CHO834-2, pp. 3B-1--3B-12.

Decell, H.P. and Smiley, W. "Householder Transformations and Optimal Linear Combinations". Report # 38 NAS-9-12777, Dept. of Mathematics, Univ. of Houston, Texas.

Householder, Alston S. (1958). "Unitary Triangularization of a Non-Symmetric Matrix". J. Assoc. Comput. Mech., 339-342.

Kaileth, T. (Feb. 1967). "The Divergence and Bhattacharyya Distance Measures in Signal Selection". IEEE Transaction on Communications Theory, Vol. 15, NO. 1, pp 52-60.

Kullback, Solomon (1968). Information Theory and Statistics. New York: Dover Publications.