IN-61- CR
46422
P- 15

Classification of Posture Maintenance Data with Fuzzy Clustering Algorithms

Interim Progress Report

James C. Bezdek
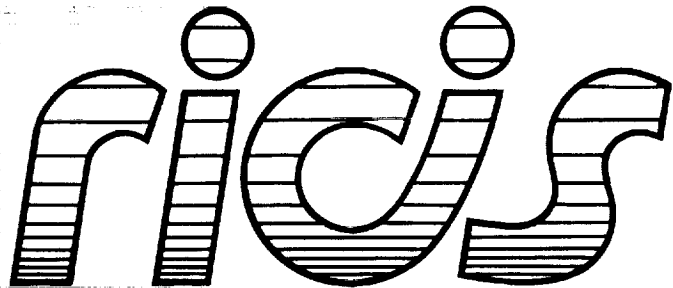
August 15, 1991

Research Institute for Computing and Information Systems
University of Houston - Clear Lake

T·E·C·H·N·I·C·A·L   R·E·P·O·R·T

The
RICIS
Concept

The University of Houston-Clear Lake established the Research Institute for Computing and Information systems in 1986 to encourage NASA Johnson Space Center and local industry to actively support research in the computing and information sciences. As part of this endeavor, UH-Clear Lake proposed a partnership with JSC to jointly define and manage an integrated program of research in advanced data processing technology needed for JSC's main missions, including administrative, engineering and science responsibilities. JSC agreed and entered into a three-year cooperative agreement with UH-Clear Lake beginning in May, 1986, to jointly plan and execute such research through RICIS. Additionally, under Cooperative Agreement NCC 9-16, computing and educational facilities are shared by the two institutions to conduct the research.

The mission of RICIS is to conduct, coordinate and disseminate research on computing and information systems among researchers, sponsors and users from UH-Clear Lake, NASA/JSC, and other research organizations. Within UH-Clear Lake, the mission is being implemented through interdisciplinary involvement of faculty and students from each of the four schools: Business, Education, Human Sciences and Humanities, and Natural and Applied Sciences.

Other research organizations are involved via the "gateway" concept. UH-Clear Lake establishes relationships with other universities and research organizations, having common research interests, to provide additional sources of expertise to conduct needed research.

A major role of RICIS is to find the best match of sponsors, researchers and research objectives to advance knowledge in the computing and information sciences. Working jointly with NASA/JSC, RICIS advises on research needs, recommends principals for conducting the research, provides technical and administrative support to coordinate the research, and integrates technical results into the cooperative goals of UH-Clear Lake and NASA/JSC.

Preface

This research was conducted under auspices of the Research Institute for Computing and Information Systems by Dr. James C. Bezdek of the Institute for Interdisciplinary Study of Human and Machine Cognition at the University of West Florida. Dr. Terry Feagin served as RICIS research coordinator.

The views and conclusions contained in this report are those of the author and should not be interpreted as representative of the official policies, either express or implied, of NASA or the United States Government.

CLASSIFICATION OF POSTURE MAINTENANCE DATA

WITH FUZZY CLUSTERING ALGORITHMS

INTERIM PROGRESS REPORT

Submitted to :

NASA Johnson Space Center
c/o Dr. James Villereal, Director
Neural Networks Program
Software Technology Branch, JT4
Houston, Texas 77058

Submitted by :

The University of West Florida
Institute for Interdisciplinary Study of Human and Machine Cognition
Division of Computer Science
11000 University Parkway
Pensacola, FL 32514

Date Submitted : August 15, 1991

Principal
Investigator :

Date : 8/11/91

Dr. James C. Bezdek : Nystul Professor : Division of Computer Science
University of West Florida, Pensacola, FLA 32514 : Tel. (904) 474-2784

Faculty
Collaborator:

Date :

Dr. Patrick O. Bobbie : Assoc. Professor : Division of Computer Science
University of West Florida, Pensacola, FLA 32514 : Tel. (904) 474-2784

Table of Contents

Executive Summary

Sensory inputs from the visual, vestibular and proprioreceptive systems are integrated by the central nervous system to maintain postural equilibrium. Sustained exposure to microgravity causes neurosensory adaptation during spaceflight, which results in decreased postural stability until readaptation occurs upon return to the terrestrial environment. Data which simulate sensory inputs under various conditions have been collected in conjunction with Johnson Space Center postural control studies using a tilt-translation device. The University of West Florida proposed applying the fuzzy c-means clustering algorithms to this data with a view towards identifying various states and stages.

Data supplied by NASA/JSC via Tom Collins, Krug Life Sciences, were submitted to the Fuzzy c-Means (FCM) algorithms in an attempt to identify and characterize cluster substructure in a mixed ensemble of pre and post adaptational TTD data. Following several unsuccessful trials with FCM using a full 11 dimensional data set, we discovered a set of two channels (features) enables FCM to separate pre from post. Our main conclusions are that FCM seems able to separate pre from post TTD subject #2 on the one trial we have used so far, but only in certain subintervals of time; and that Channels 2 (right rear transducer force) and 8 (hip sway bar) contain better discrimination information than other supersets and combinations of the data we have tried so far.

Subsequently, we have been studying separability of pre and post across the time axis, by subdividing the data into "time slices" or subintervals of the 20 second data collection epoch. So far, our tests have involved only one subject and one TTD test. For this data, several measures of separability, as well as the error rate computed by resubstitution of the training data, suggest the FCM can separate certain time slices much more readily than others. We will continue to widen our computational experience, by enlarging the tests we have conducted across different trials on the same subject, and then different subjects.

Fuzzy c-Means

Let (c) be an integer, $1 < c < n$ and let $X = \{x_1, x_2, ..., x_n\}$ denote a set of (n) feature vectors in $\mathcal{R}^p$. X is *numerical object data*; the j-th object in this study is a set of p measurements of sensor signals at time t. To be technically accurate, the notation for the posture control data should be something like $x_j = x(t_j)$, $j = 1, 2$, ... , n; however, in the interests of clarity we will suppress the dependency of the feature vectors on time. x_{jk} is, for this data, the j-th channel value associated with time k. Given X, we say that (c) fuzzy subsets $\{u_i : X \rightarrow [0,1]\}$ are a fuzzy c-partition of X in case the (cn) values $\{u_{ik} = u_i(x_k), 1 \leq k \leq n, 1 \leq i \leq c\}$ satisfy three conditions:

$$0 \leq u_{ik} \leq 1 \text{ for all } i,k ; \qquad\qquad\qquad\qquad (1a)$$

$$\Sigma u_{ik} = 1 \text{ for all } k ; \qquad \text{and} \qquad\qquad\qquad (1b)$$

$$0 < \Sigma u_{ik} < n \text{ for all } i. \qquad\qquad\qquad\qquad (1c)$$

Each set of (cn) values satisfying conditions (1) can be arrayed as a (cxn) matrix $U = [u_{ik}]$. The set of all such matrices are the *non-degenerate fuzzy c-partitions* of X:

$$M_{fcn} = \{U \text{ in } \mathcal{R}^{cn} \mid u_{ik} \text{ satisfies conditions (1) for all i and k}\}. \qquad (2)$$

And in case all the u_{ik}'s are either 0 or 1, we have the subset of *hard (or crisp) c-partitions* of X:

$$M_{cn} = \{U \text{ in } M_{fcn} \mid u_{ik} = 0 \text{ or } 1 \text{ for all i and k}\}. \qquad\qquad (3)$$

Data structures identified by partitions which are optimal in the sense of minimizing the function defining them often provide good insights and explanations into substructure of the process that produced the data. The FCM functional is as follows:

$$J_m(U,v;X) = \Sigma\Sigma u_{ik}{}^m(\|x_k\text{-}v_i\|_A)^2 \quad , \qquad \text{where} \tag{5}$$

$m \in [1, \infty)$ is a weighting exponent on each fuzzy membership; $\qquad$ (6a)

$U \in M_{fcn}$ is a fuzzy c-partition of X; $\qquad$ (6b)

$v = (v_1, v_2, ..., v_c)$ are cluster centers in $\mathfrak{R}^s$; $\qquad$ (6c)

A = is any positive definite (s x s) matrix; $\qquad$ and $\qquad$ (6d)

$\|x_k\text{-}v_i\|_A = (x_k\text{-}v_i)^T A (x_k\text{-}v_i)$ is the OG distance (in the A norm) from x_k to v_i . $\qquad$ (6e)

Conditions necessary for a local minimum of J_m are as follows:

Fuzzy c-Means (FCM) Theorem [4]. (U,v) may minimize $\Sigma\Sigma u_{ik}{}^m(\|x_k\text{-}v_i\|_A)^2$ for m>1 only if :

$$u_{ik} = \quad (\Sigma \|x_k\text{-}v_i\|_A \text{ / } \|x_k\text{-}v_j\|_A)^{-2/(m-1)} \qquad \text{for all } i,k \qquad ; \qquad \text{and} \tag{7a}$$

$$v_i \quad = \Sigma(u_{ik})^m x_k / \Sigma(u_{ik})^m \qquad\qquad \text{for all } i \quad . \tag{7b}$$

The FCM algorithms are simple Picard iteration through (8) :

Fuzzy/Hard c-Means (FCM/HCM) Algorithms [2].

<FCM/HCM 1> : Given unlabeled data set $X = \{x_1, x_2, ..., x_n\}$. Fix : $1 \leq c < n$; $1 < m < \infty$ (/ m=1 for HCM); positive definite weight matrix A to induce an inner product norm on $\mathfrak{R}^s$; and ϵ, a small positive constant.

<FCM/HCM 2>: Guess $v_0 = (v_{1,0} , v_{2,0} , ..., v_{c,0}) \in \mathfrak{R}^{cs}$ $\qquad$ (or, initialize $U_0 \in M_{fcn}$).

<FCM/HCM 3>: For j = 1 to J:
$\qquad$ <3a> : Calculate U_j with $\{v_{i,j-1}\}$ and $\qquad$ (8a) ;
$\qquad$ <3b>: Update $v_{i,j-1}$ to $v_{i,j}$ with U_j and (8b), $1 \leq i \leq c$
$\qquad$ <3c>: If max{ $\|v_{i,j-1} - v_{i,j}\|$ }$\leq \epsilon$, then stop and put $(U^*,v^*) = (U_j,v_j)$; Else : Next j

Configuration of the Posture Control Data

The following conceptual arrangement of the data will be referred to in subsequent discussion. We regard the data as an array of size (p x 4000), where p=number of features (channels) used in the processing. Each column of the matrix is thus a vector in $\mathfrak{R}^p$; and each row of the matrix contains the observations collected by one sensor at each point in time. Further, the data are labeled as pre and post, so the overall data matrix is partitioned at column 2000 (the final observation time). EMG data were sampled at four times the frequency of transducer data, so we decimated the EMG data in order to align them with the transducer samples.

The basic data set then consists of 2000 samples taken across a 20 second time interval with sensors attached to a subject at 11 locations (channels). Data were collected both before (pre) and after (post) a subject was exposed to roughly 30 minutes in the TTD with one of six trial environments (Trials 1-6). When using FCM, rows of the data matrix X in Figure 1 correspond to features. For p=11, all of the data channels are used. Choosing, e.g., features 2,5, and 8 corresponds to reading and processing only those three rows of X. The vector $x_{pre,1}$ which is highlighted in Figure 1 is a column vector with p entries : $x_{pre,1}$ = $(x_{pre,1,1}, x_{pre,1,2}, \ldots, x_{pre,1,p})^T$. It will be convenient in our discussion to subscript data sets as follows:

$$X_{ij} = \text{data matrix for subject i, trial j} \qquad 1 \leq i \leq 5 \text{ and } 1 \leq j \leq 6. \qquad (8)$$

Conceptually, the data matrix has the following configuration:

Figure 1. Arrangement of the Posture Control Data for one subject for one trial
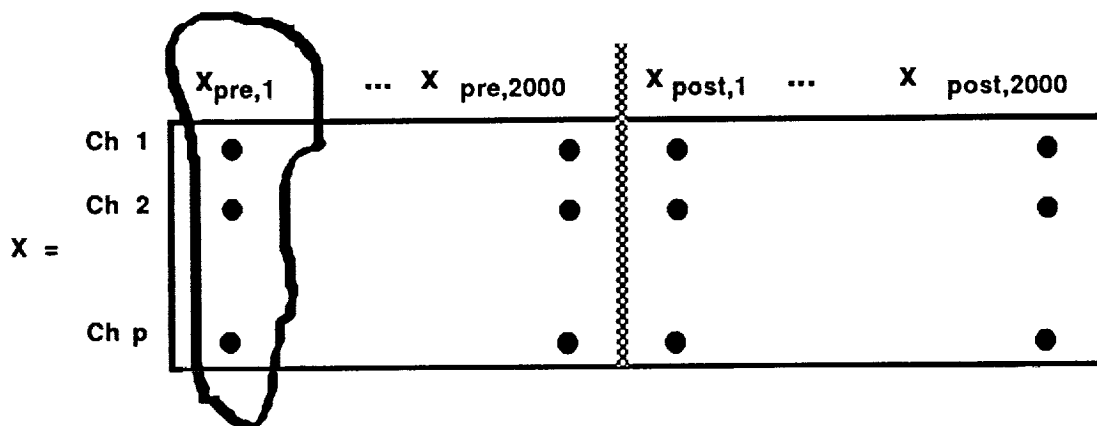
The 11 features in X are labeled as shown in Table 1 (Feature # F = NASA Channel # C):

Table 1. Posture Control Features

F = C	Location
1 = 1	left front transducer force
2 = 2	right front transducer force
3 = 3	shear force transducer
4 = 4	left rear force transducer
5 = 5	right front force transducer
6 = 7	shoulder sway bar
7 = 8	hip sway bar
8 = 11	soleus
9 = 12	hamstrings
10 = 13	tibialis
11 = 14	quadriceps

Feature Selection

After several runs using all 11 channels, each of which produced uninterpretable results, we performed several statistical analyses (principle components and MANOVA) in an attempt to find transformations of the data that would give better results in 11 space. These attempts were also short lived, and seemed to produce nothing useful. Finally, we resorted to a graphical plot of the raw signals in all 11 channels, and used visual inspection to select the signal channels that seemed most likely to possess good discriminatory power. The features (channels) selected for further analysis were as follows:

Feature Set 1
Channel 2 = right rear force transducer
Channel 5 = right front force transducer
Channel 8 = hip sway bar

None of the EMG data seemed, upon visual inspection at least, to contain information that could be used to elicit classification, so we abandoned processing on these channels early in the study. At the suggestion of Tom Collins, we also tried the following sets of three features:

Feature Set 2
Channels (1+2+4+5)/4 = ave. left, right,front,rear force transducers
Channel 3 = shear force transducer
Channel 8 = hip sway bar

Feature Set 3
Channels (1+2+4+5)/4= ave. left, right,front,rear force transducers
Channel 3 = shear force transducer
Channel 7 = shoulder sway bar

However, neither of these feature sets seemed to produce better results than the channel 3-tuple {2,5,8}. We will continue to experiment with Feature sets 1, 2 and 3 as our computational base widens to include further trials and subjects.

Initialization of FCM for the Posture Control Data

Since X is labeled (pre or post), we can initialize FCM in step FCM 2 with U_L, the hard partition that labels the data. Moreover, the number of classes is known, c=2. Thus, partition U_L is the 2 x 4000 matrix :

$$U_L = \begin{vmatrix} 1\,1\,1 \dots\dots\dots 1\,1 & 0\,0\,0 \dots\dots\dots 0\,0 \\ 0\,0\,0 \dots\dots\dots 0\,0 & 1\,1\,1 \dots\dots\dots 1\,1 \end{vmatrix} \begin{matrix} \blacktriangleright\text{ pre} \\ \\ \blacktriangleright\text{ post} \end{matrix} \quad (9)$$

This initialization cannot be used, of course, with unlabeled data, so initialization procedures for FCM will have to be widened as the study progresses. For calculations on time subintervals, a label matrix in the form of (10), adjusted to the correct subsize, is used to initialize FCM, and is the basis for computation of the resubstitution error rate described next.

Measures of Performance and Separability

We use three performance indices to guide our analysis of the data. The primary measure of performance is the <u>observed label error rate</u> $E_L(U, X_{ij})$ for U in M_{cn}. This is computed by first converting any terminal fuzzy c-means partition, say U_{FCM}, into a hard partition by thresholding with the so-called method of α-cuts. Specifically, for a chosen membership threshold $\alpha \in [0,1]$, we define the hard c-partition U_α derived from U_{FCM} as follows:

For cols j for which there is a row i in UFCM such that $U_{FCM,ij} \geq \alpha$, $u_{\alpha,ij} = 1$, $u_{\alpha,ij} = 0$, k ≠ i; and otherwise, For cols j for which there is no row i in UFCM such that $u_{FCM,ij} \geq \alpha$, declare "no label for j" .

Because "no label for j" columns of U_α do not contain a "1" in any row, U_α is not, strictly speaking a hard partition of the data. This can be accounted for in a formal way by adding a c+1-st row to U_α and U_L, with zeroes in every column of U_L, and (placed) 1's in each column of U_α where "no label" occurs. After the hard "partition" U_α has been determined, we compute the label error rate as follows:

$$E_L(U_{FCM}, X_{ij}) = \Sigma\Sigma |u_{L,ij} - u_{\alpha,ij}| / 2n_L \quad (10)$$

where n_L is the number of labeled data used for the run. E_L is simply the number of times that the labels in U_α disagree with the given labels divided by the total number of trials (samples) used to generate U_{FCM}.

We are also using two measures of separability of the data that are related to E_L, and are thus most accurately regarded as"second order"measures of classifier performance. The reason we are studying these measures is to find a means for detecting, in <u>unlabeled</u> data, when the data are being well separated, since the error rate E_L cannot be computed with unlabeled data in on-line processing during dat acquisition. These measures are as follows:

Cluster Center Separation (c=2, Euclidean Norm,)

$$DV(v_{FCM}) = ||\, v_{FCM,\, pre} - v_{FCM,\, post}\, || \qquad\qquad (11)$$

It is intuitively plausible, but not mathematically necessary, that DV increase as the clusters that have $v_{FCM,\, pre}$ and $v_{FCM,\, post}$ as their prototypes become increasingly well separated. This is illustrated schematically in Figure 2:

Figure 2. Geometric Rationale for the measure of Cluster Center Separation
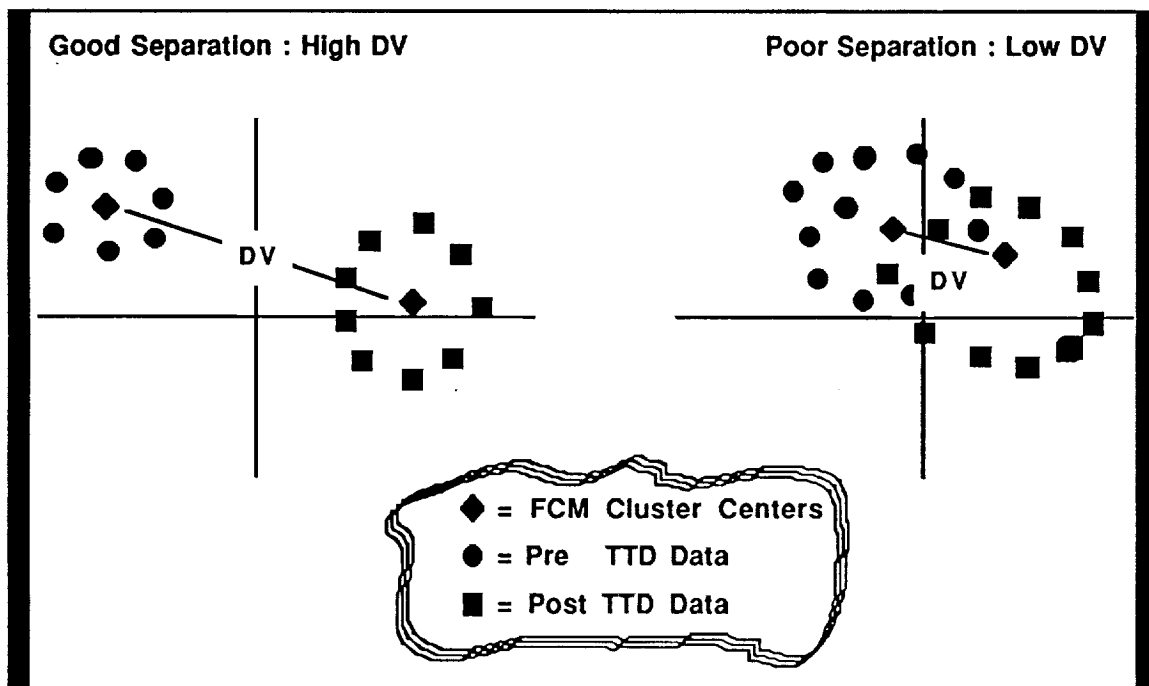
Partition Entropy

$$H(U_{FCM}) = - \Sigma\Sigma u_{ik} \ln_e (u_{ik})/n_L \qquad (12)$$

This index is simply one measure of "fuzziness" of any partition in M_{fcn}. The value of H is 0 on every hard U in M_{cn}, and increases to a maximum of $\ln_e(c) = 0.69 ...$ for c=2 as U approaches the "fuzziest" partition, namely, $u_{ij} = (1/c)$ for all i and j. Thus, the better separated the clusters, the harder is U_{FCM}, and the closer $H(U_{FCM})$ is to zero. H is often (but not always!) a good indication of (algorithmic) separability of X, so we will track its ability to indicate when FCM is doing well.
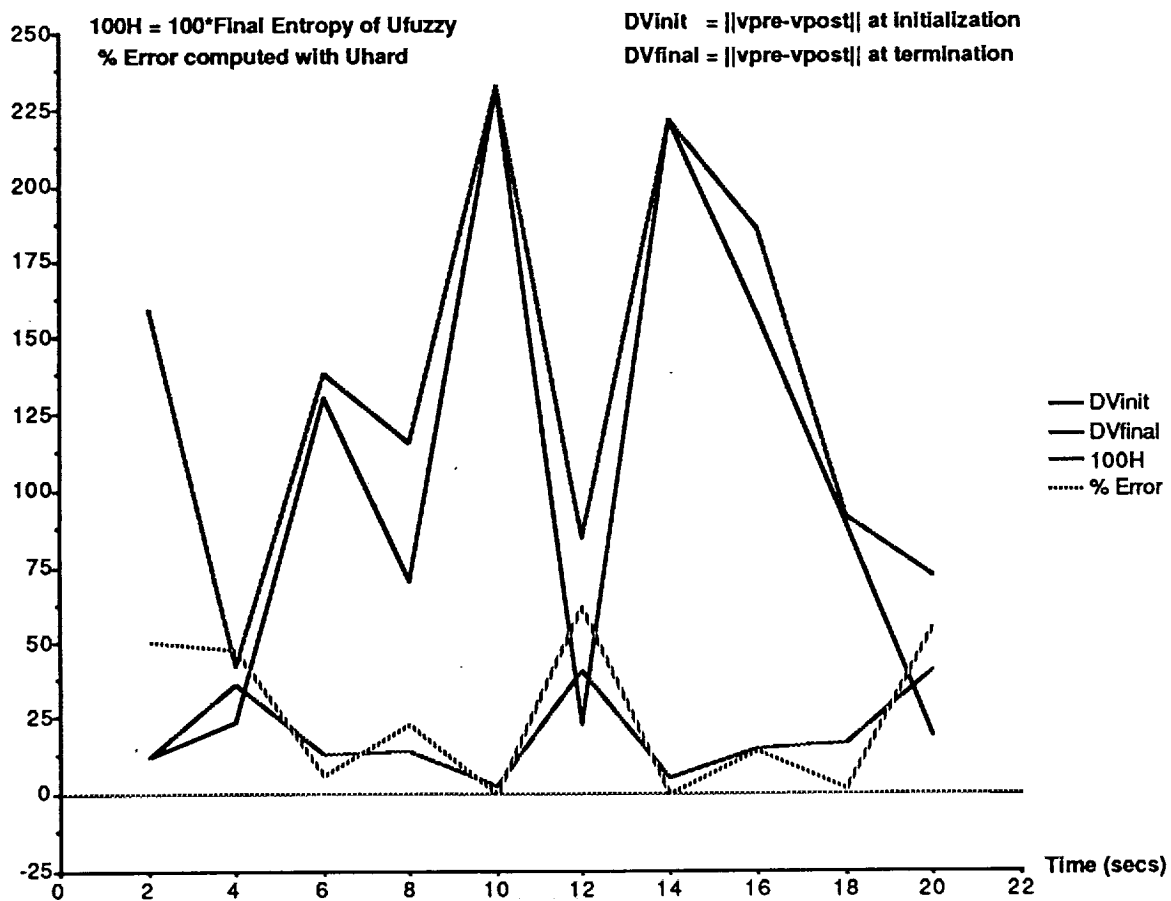
Initial Results and Discussion

Several runs using Feature set #1 = Channels {2,5, 8} and various 2 dimensional subsets thereof were made on all 4000 data vectors. These runs were notable only in their lack of success. Following this, we began subdividing X into time slices, and processing subinterval data sets. That is, we took a vertical subslice through the matrix X in Figure 1, adjusted U_0 and n_L, and submitted the data to FCM. This has been done over a number of different feature subsets and time slices. The most striking results are illustrated in Table 2 and Figure 3, which is a plot of the data in Table 2.

Table 2. Performance Indices for data X_{22} on time slices of 2 seconds each

T(secs)	DVInit $DV(v_{FCM,init})$	DVfinal $DV(v_{FCM,final})$	100*Entropy $100^* H(U_{FCM})$	% Error $100^* E_L(U_{FCM}, X_{ij})$
2	12	159.36	12	50
4	24	42.83	36	47.5
6	130	137.88	13	6
8	70	115.51	14	22.25
10	232	232.63	3	0
12	23	84.32	40	61.5
14	221	221.21	5	0
16	158	185.64	15	14.25
18	89	91.21	17	1.75
20	20	71.9	40	55

Figure 3. Performance Indices for data X_{22} on time slices of 2 seconds (cf. Table 2)

100H = 100*Final Entropy of Ufuzzy
% Error computed with Uhard

DVinit = ||vpre-vpost|| at initialization
DVfinal = ||vpre-vpost|| at termination

— DVinit
— DVfinal
— 100H
······ % Error

Time (secs)

Note that the distance DV between cluster centers was computed twice; at initialization, using (7b) with U_L; and again at termination, using (7b) with U_{FCM}. Figure 3 shows that DV increases and H decreases whenever E_L, the labeling error rate, is close to zero. The tentative inference to be drawn from this graph is that FCM is able to differentiate between pre and post signals at some subintervals of the data collection epoch. These results suggest that either DV or H or both may be indicators of the error rate across such time steps (remember, there is no way to know what the error rate is with truly unlabeled data). However, one must view this conjecture with great caution; it would be foolish to impute much credence to this hypothesis on the basis of outputs from one subject on one trial. Our goal for the remainder of the project will be to wider our computational experience across both trials on the same subject, as well as across different subjects. Perhaps this initial conjecture will not stand up across different data sets; on the other hand, we feel that these initial results are at least encouraging.

4. References

[1] Kohonen, T. Self-Organization and Associative Memory, 3rd Edition, Springer-Verlag, Berlin, 1989.

[2] Bezdek, J. Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, New York, 1981.

[3] Duda, R. and Hart, P. Pattern Classification and Scene Analysis, Wiley, New York, 1973.

[4] Tou , J. and Gonzalez, R. Pattern Recognition Principles, Addison-Wesley, Reading, 1974.

[5] Hartigan, J. Clustering Algorithms, Wiley, New York, 1975.

[6] Pao, Y.H. Adaptive Pattern Recognition and Neural Networks, Addison-Wesley, Reading, 1989.

[7] Lippman, R. An Introduction to Neural Computing, IEEE ASSP Magazine, April, 1987, 4-22.

[8] Palosky, W., Harm, D., Reschke, M., Doxey, D., Skinner, N., Michaud, L. and Parker, D. Postural Changes following Sensory Reinterpretation as an Analog to Spaceflight, private communication, 1990.

[9] Reschke, M. Neural Network Modeling of Postural Control, NASA/JSC technical report, 1990.

[10] Bezdek, J. A Note on Generalized Self-organizing clustering Algorithms, in Proc. SPIE Applications of AI (8), ed. M. Trivedi, V1293, 1990, 260-267.

[11] Huntsberger, T. and Ajjimarangsee, P. Parallel Self-Organizing Feature Maps for Unsupervised Pattern Recognition, in press, Int'l. Jo. General Systems, 1989.

[12] Dubes, R. How Many Clusters are Best? An Experiment, Patt. Recog., 20, 1987, 645-663.

[13] Ball, G. and Hall, D. A Clustering Technique for Summarizing Multivariate Data, Behav. Sci., 12, 1967, 153-155.

[14] Dunn, J.C. A Fuzzy Relative of the ISODATA Process and its use in Detecting Compact, Well-Separated Clusters, Jo. Cybernetics, 3, 1974, 32-57.