NRL Memorandum Report 3407

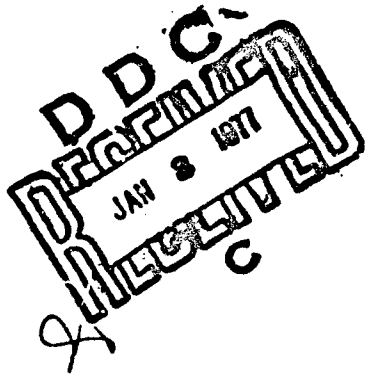
# Application of Automatic Clustering to Emitter Identification

JAMES SLAGLE AND RICHARD C. T. LEE

*Computer Science Laboratory*
*Communications Sciences Division*



November 1976





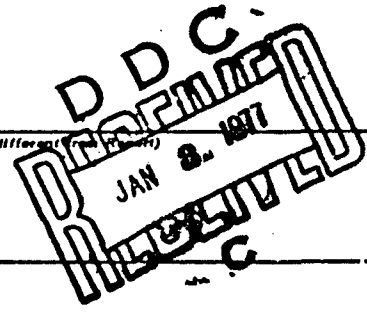
NAVAL RESEARCH LABORATORY
Washington, D.C.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|
| 1 REPORT NUMBER: NRL Memorandum Report 3407 | 2 GOVT ACCESSION NO. / 3 RECIPIENT'S CATALOG NUMBER |
| 4 TITLE (and Subtitle): APPLICATION OF AUTOMATIC CLUSTERING TO EMITTER IDENTIFICATION | 5 TYPE OF REPORT & PERIOD COVERED: An interim report on a continuing NRL problem. |
| | 6 PERFORMING ORG. REPORT NUMBER |
| 7 AUTHOR(s): James Slagle and Richard C.T. Lee | 8 CONTRACT OR GRANT NUMBER(s) |
| 9 PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Research Laboratory, Washington, D.C. 20375 | 10 PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NRL Problem B02-23 |
| 11 CONTROLLING OFFICE NAME AND ADDRESS: Naval Electronic Systems Command, Washington, D.C. 20360 | 12 REPORT DATE: November 1976 |
| | 13 NUMBER OF PAGES: 12 |
| 14 MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | 15 SECURITY CLASS. (of this report): UNCLASSIFIED |
| | 15a DECLASSIFICATION/DOWNGRADING SCHEDULE |

16 DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.




17 DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18 SUPPLEMENTARY NOTES

19 KEY WORDS (Continue on reverse side if necessary and identify by block number)

Clustering
Unsupervised learning
Pattern recognition
Nearest neighbor
Algorithm
Feature selection algorithm

20 ABSTRACT (Continue on reverse side if necessary and identify by block number)

The goal of clustering is the partitioning of a given set of objects into subsets called clusters in such a way that the objects in a cluster are similar to one another and that objects in different clusters are dissimilar. Clustering may help in getting a more or less direct understanding of the relationships among the objects, and it may be useful as a first step in pattern recognition. Some possible applications are automatic phoneme recognition, data base management systems, personnel classification, detection of errors in files and computer security.

(Continues)

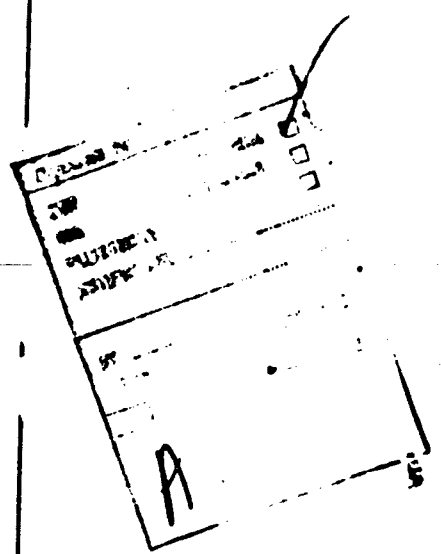
DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

251 150

20. Abstract (Continued)

Several clustering methods were applied to data sets of practical importance. Automatic pattern recognition using the k nearest neighbors was applied. An efficient method for selecting a good subset from the full set of 44 features was tried. In all cases, the results were good.



CONTENTS

# APPLICATION OF AUTOMATIC CLUSTERING TO EMITTER IDENTIFICATION

## 1. INTRODUCTION

Research in and applications of artificial intelligence [4] are the main thrusts of the Computer Science Laboratory, Communications Sciences Division, Naval Research Laboratory. At present, we are working on phoneme recognition for a low band width speech communication system, a computer-controlled manipulator, and intelligent data base management system, as well as automatic clustering and its application to emitter identification. In this paper, we shall focus on the automatic clustering work. The goal of clustering [1,2,3,5,6] is the partioning of a given set of objects into subsets called clusters in such a way that the objects in a cluster are similar to one another and that objects in different clusters are dissimilar. Two objects may be considered similar if, for example, the euclidean distance between them in the measurement (feature) space is small.

Clustering has two main purposes. First, clustering may help in getting a more or less direct understanding of the relationships among the objects. Second, clustering may be useful as a first step in pattern recognition. In pattern recognition (unlike clustering), each presented object must be labeled with its class membership. After clustering (with labels ignored), one can "look at" the data to estimate whether pattern recognition will be easy or difficult, whether a given class should be combined with another class or divided into two or more classes, etc.

Since clustering is quite general, it has many applications. Later we shall see that clustering can be applied to emitter identification. It can be applied to automatic phoneme recognition and personnel classification. It is often very important to find a cluster having only one member. Such a cluster may be a mistake in the data base or represent a very unusual object. An unusual use of a computer system may be an attempted security penetration.

## 2. CLUSTERING THE EMITTER IDENTIFICATION DATA

Each row in Table 1 is a sample (object) representing 18 measurements of an emission. Our sponsor gave us 18 samples. Without knowing anything else about the data, we applied several clustering techniques, some of which we had developed [2,5,6]. We obtained the following four clusters:

(1) 1, 2
(2) 3, 5, 6, 7, 9, 10, 11
(3) 4, 17, 18
(4) 8, 12, 13, 14, 15, 16

This means that samples (rows) 1 and 2 are in cluster (1), etc.

Note: Manuscript submitted October 29, 1976.

Table 1. Original Data for Emitter Identification

| EMISSIONS | PARAMETERS 1 | 2 | 3 | 4 | 5 | 6 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | - .10 | - .07 | -1.26 | - .76 | .74 | 2.47 | - .46 | - .44 | - .97 | - .50 | 0.00 | .14 |
| 2 | - .33 | - .40 | -1.02 | - .76 | .12 | 1.95 | - .44 | - .27 | -1.80 | -1.00 | - .53 | - .62 |
| 3 | 1.01 | 1.23 | 1.18 | .85 | -1.17 | .26 | .85 | 1.70 | 1.60 | .50 | .36 | .38 |
| 4 | - .99 | -1.05 | 1.31 | .85 | - .37 | - .75 | - .46 | - .56 | .32 | - .33 | - .79 | .83 |
| 5 | -1.24 | 1.23 | .71 | .85 | - .71 | .22 | - .15 | -1.05 | 1.00 | .83 | .64 | .31 |
| 6 | 1.02 | .90 | .35 | .44 | - .61 | .49 | - .80 | .19 | - .36 | 2.50 | 3.36 | .90 |
| 7 | 1.57 | 1.55 | .65 | .85 | - .60 | .21 | - .18 | - .98 | 1.03 | -1.50 | - .64 | .13 |
| 8 | - .49 | .72 | - .96 | -1.56 | .27 | - .65 | - .40 | .28 | -1.58 | - .33 | 0.00 | .23 |
| 9 | .13 | - .40 | .61 | .45 | -1.43 | .91 | - .62 | -1.34 | 1.23 | .67 | .64 | .33 |
| 10 | 1.48 | 1.55 | .71 | .85 | 3.19 | - .70 | .82 | .42 | 1.16 | .33 | .71 | .36 |
| 11 | .53 | .58 | .28 | .45 | - .40 | .50 | .78 | .37 | .36 | 1.33 | .64 | .43 |
| 12 | .25 | - .40 | - .97 | -1.16 | - .33 | - .95 | - .39 | .20 | - .97 | 1.67 | - .64 | - .52 |
| 13 | - .61 | - .72 | 1.05 | -1.16 | .48 | - .98 | - .37 | - .88 | .58 | - .33 | -1.43 | - .62 |
| 14 | 1.48 | -1.37 | -1.13 | - .76 | .66 | - .69 | 3.43 | - .19 | - .39 | .17 | .36 | .27 |
| 15 | .03 | .25 | .94 | - .76 | .45 | - .24 | - .37 | - .14 | - .29 | -1.50 | .64 | .27 |
| 16 | -1.85 | -1.70 | - .08 | -1.16 | .50 | - .23 | - .37 | 1.13 | 1.03 | - .50 | - .29 | .18 |
| 17 | - .61 | - .40 | 1.47 | 1.65 | - .46 | - .67 | - .54 | .21 | -1.52 | .17 | .21 | .29 |
| 18 | - .94 | - .72 | 1.01 | 1.25 | .17 | -1.13 | - .51 | 2.89 | .55 | - .17 | - .50 | .11 |

Our sponsor then told us that the "real" clusters should be as follows:

(1') 5, 6, 10, 11

(2') 4, 17, 18

(3') 3, 7, 9

(4') 1, 2, 8, 12, 13, 14, 15, 16

Thus, the results were quite good. Careful analysis of the data revealed the following:

(a) If cluster (1) has to be combined with any other cluster, it should be combined with cluster (4). However, cluster (1) is definitely different from cluster (4).

(b) From only the given data, one really cannot say that cluster (1') and cluster (3') should be two clusters rather than being combined into one cluster.

Our sponsor agreed with both of these statements.

We were then supplied with 12 new samples (rows) with some missing values. (See Table 2). After using various clustering methods, we concluded that A, B, C, D, E, and F belong to cluster (1'), that G, H, and I belong to (2'), and that J, K, and L belong to either (1') or (3'). Then our sponsor told us that the first two conclusions are absolutely correct. He said that samples J, K, and L belong to cluster (3'). We could say only that they belong to either (1') or (3'), because (1') and (3') are so close to each other.

## 3. TWO CLUSTERING METHODS

Two of the six clustering methods we used are data reorganization [5] and two principal components [1,3]. The other four are minimal spanning trees [1,3], non-linear mapping [6], and two versions of a triangulation method [2]. We do not describe these here, since they are fairly complicated to explain and are described elsewhere. In data reorganization, rows (and sometimes columns as here) are permuted to put similar rows (and columns) together. (Two n-tuples are similar to the extent that the n-dimensional distance between them is small.) The result for the 18 samples is shown in Table 3. After this reorganization was done, the sponsor told us the classes. Table 3 is perfect in the sense that it could be broken between rows into four parts, each exactly corresponding to a given class. We also used the reorganization by rows for automatically obtaining the clustering hierarchy shown in Fig. 1. Table 3 is broken into two tables by dividing between the rows whose distance apart is largest. These tables are divided further, etc. Partly on this hierarchy, we obtained the clusters (1) through (4) given earlier.

Table 2. Emitter Identification Data With Missing Values

| EMISSIONS | PARAMETERS | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 1.65 | 1.55 | .92 | .42 | .85 | - .23 | 1.62 | .78 | — | — | .33 | .21 | .28 |
| B | 1.70 | 1.55 | .66 | .44 | — | — | — | - .16 | — | — | — | — | — |
| C | 1.19 | 1.23 | .65 | .44 | - .81 | .30 | 1.39 | .76 | - .82 | 1.23 | 1.50 | .86 | — |
| D | 1.15 | 1.23 | .91 | 1.25 | - .81 | - .40 | — | - .12 | — | — | — | — | — |
| E | 1.65 | 1.55 | .68 | .85 | -1.28 | .35 | — | .82 | — | — | — | — | — |
| F | 1.43 | 1.23 | .59 | .45 | -1.19 | .40 | .43 | - .20 | 2.32 | — | — | — | — |
| G | - .22 | - .40 | 1.60 | 1.25 | 0.58 | - .82 | — | - .62 | — | — | — | — | — |
| H | -1.37 | -1.37 | 1.29 | 1.65 | .26 | - .70 | — | .69 | — | — | — | — | — |
| I | - .18 | - .40 | 1.71 | 1.25 | — | — | — | .70 | — | — | — | — | — |
| J | 1.32 | 1.23 | 1.08 | 1.25 | - .78 | .49 | — | - .15 | — | — | — | — | — |
| K | 1.41 | .90 | .99 | .85 | — | — | — | .65 | — | — | — | — | — |
| L | .46 | .25 | .93 | .45 | — | — | — | - .20 | — | — | — | — | — |

Table 3. Data Reorganized by Row and Column

| | 14 | 18 | 17 | 15 | 3 | 9 | 4 | 15 | 7 | 8 | 1 | 2 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.44 | 0.14 | 0.00 | -0.50 | -1.26 | -1.04 | -0.76 | -0.97 | -1.06 | 0.00 | -0.10 | -0.07 | 1.22 |
| 2 | 0.27 | -0.62 | -0.43 | -1.00 | -1.02 | -1.04 | -0.76 | -1.80 | -0.51 | 0.00 | -0.33 | -0.40 | 0.97 |
| 12 | 0.20 | -0.52 | -0.64 | -1.67 | -0.97 | -1.04 | -1.16 | -0.97 | -0.68 | 0.00 | 0.25 | -0.40 | -1.30 |
| 8 | 0.28 | 0.23 | 0.00 | -0.33 | -0.96 | -1.04 | -1.56 | -1.58 | -0.64 | -0.97 | -0.49 | 0.72 | -1.07 |
| 14 | -0.19 | 0.26 | 0.36 | 0.17 | -1.13 | -1.04 | -0.76 | -0.84 | -0.58 | -0.97 | -1.48 | -1.37 | -1.24 |
| 16 | 1.13 | 0.18 | -0.29 | -0.50 | -0.97 | -1.04 | -1.16 | 1.03 | 0.45 | -1.94 | -1.85 | -1.70 | -0.88 |
| 13 | -0.88 | -0.62 | -1.43 | -0.33 | 1.05 | -1.04 | -1.16 | 0.58 | 0.13 | 0.00 | -0.61 | -0.72 | -1.05 |
| 15 | -0.15 | 0.27 | 0.64 | -1.50 | 0.94 | -1.04 | -0.76 | -0.29 | -0.11 | 0.00 | 0.33 | 0.25 | -0.51 |
| 10 | 0.42 | 0.35 | 0.71 | 0.33 | 0.71 | 0.57 | 0.85 | 1.16 | 1.52 | 0.97 | 1.48 | 1.55 | 1.01 |
| 6 | 0.19 | 0.90 | 3.36 | 2.50 | 0.35 | 0.57 | 0.44 | -0.36 | 1.38 | 0.97 | 1.02 | 0.90 | 1.04 |
| 11 | 0.37 | 0.43 | 0.64 | 1.33 | 0.29 | 0.57 | 0.45 | 0.36 | 1.30 | 0.97 | 0.53 | 0.58 | 0.98 |
| 5 | 1.05 | 0.31 | 0.64 | 0.83 | 0.71 | 0.57 | 0.85 | 1.00 | 1.16 | 0.97 | 1.24 | 1.23 | 1.01 |
| 7 | -0.98 | 0.13 | -0.64 | -1.50 | 0.65 | 0.57 | 0.85 | 1.03 | 1.36 | 1.94 | 1.57 | 1.55 | 1.02 |
| 3 | -1.70 | 0.38 | 0.36 | 0.50 | 1.18 | 1.11 | 0.85 | 1.16 | 0.78 | 0.97 | 1.01 | 1.23 | 0.81 |
| 9 | -1.34 | 0.33 | 0.64 | 0.67 | 0.61 | 0.57 | 0.45 | 1.23 | -0.63 | 0.00 | 0.13 | -0.40 | 0.44 |
| 4 | 0.45 | 0.83 | -0.79 | -0.33 | 1.31 | 1.11 | 0.85 | 0.32 | -1.66 | -0.97 | -0.99 | -1.05 | -1.01 |
| 17 | 0.21 | 0.29 | 0.21 | 0.17 | 1.47 | 1.65 | 1.65 | -1.52 | -1.08 | -0.97 | -0.61 | -0.40 | -0.65 |
| 18 | 2.89 | 0.11 | -0.50 | -0.17 | 1.01 | 1.11 | 1.25 | 0.55 | -0.22 | -0.97 | -0.94 | -0.72 | -0.80 |

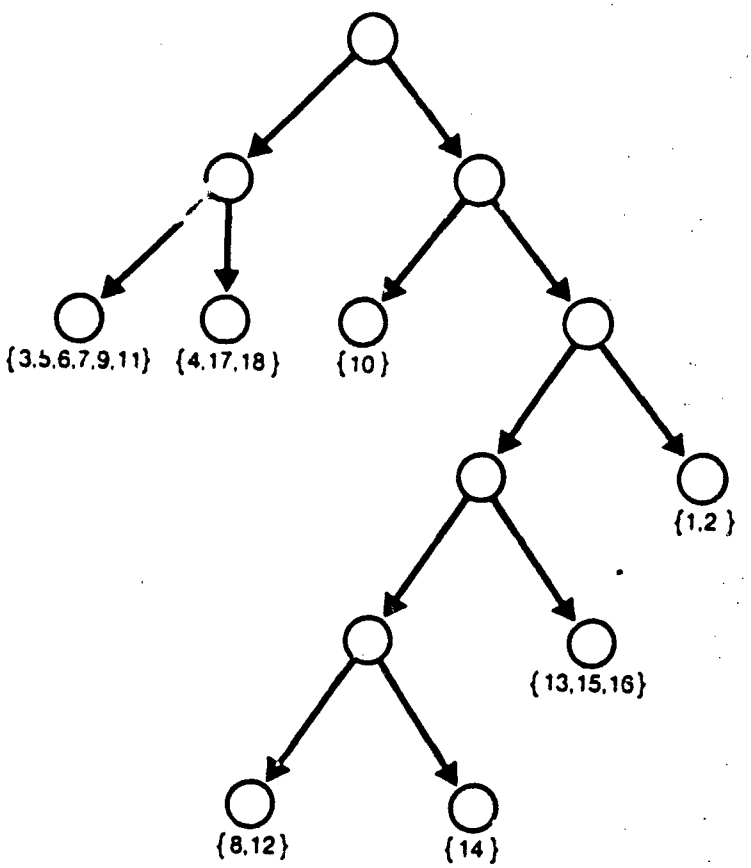



Fig. 1 — Clustering hierarchy from data reorganization by row

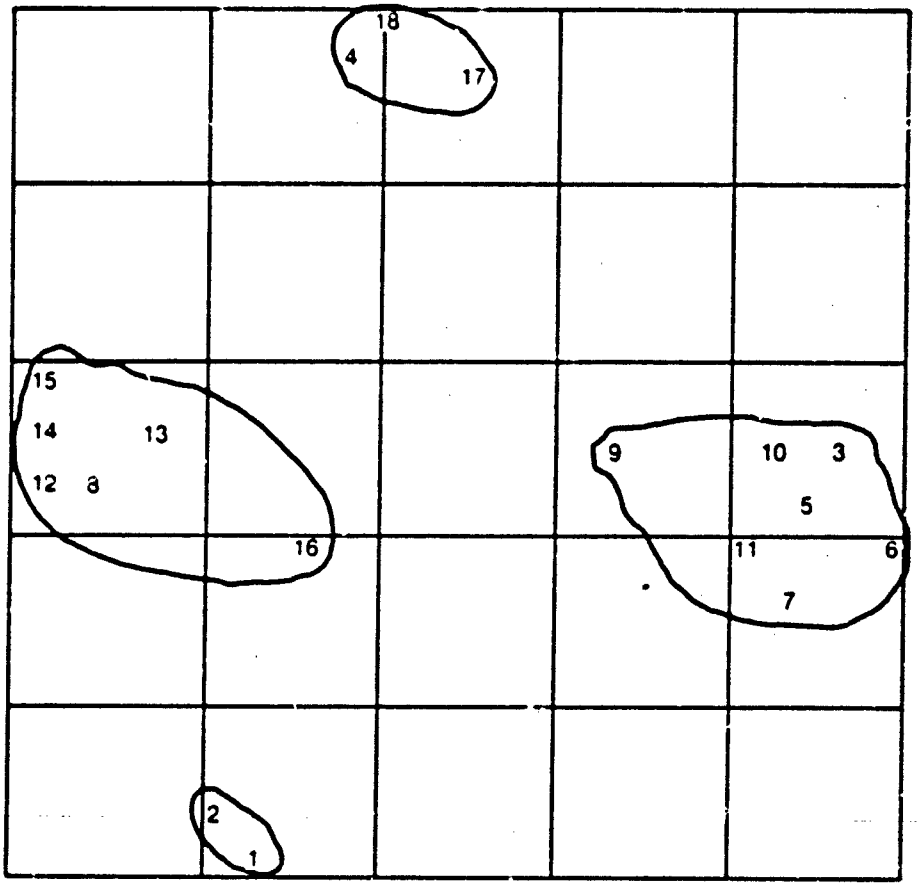



Fig. 2 — Principal components

In the method of two principal components, the data is projected from an 18 dimensional space onto the "best" plane (two dimensional space). The result for our data is shown in Fig. 2. Each of the two axes is a linear combination of the original 18 features. Again we see that the results are good. Partly based on Fig. 2, we obtained clusters (1) through (4) given earlier.

## 4. FUTURE WORK

Our sponsor will give us a much larger set of data, and we are looking forward to analyzing it. We hope that the results will continue to be favorable. We shall also be applying clustering techniques to automatic phoneme recognition and intelligent data base management systems.

## REFERENCES


[1] Duda, R. and Hart, P.: "Pattern Classification and Scene Analysis". Wiley Interscience, 1973.

[2] Lee, R. C. T., Slagle, J., and Blum, H.: "A Triangulation Method for the Sequential Mapping of Points from n-Space to 2-Space". (To be published as correspondence in IEEE Transactions on Computers.)

[3] Misel, W.: "Computer Oriented Approaches to Pattern Recognition". Academic Press, 1972.

[4] Slagle, J.: "Artificial Intelligence: The Heuristic Programming Approach." McGraw-Hill, New York, 1971. Also available in German and Japanese translations.

[5] Slagle, J., Chang, C.L., and Heller, S.: A Clustering and Data Reorganizing Algorithm, IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC-5, (Jan. 1975) pp. 125-128.

[6] Slagle, J., Chang, C.L., and Lee, R.C.T.: Experiments with some Cluster Analysis Algorithms", Pattern Recognition, Vol. 6, 1974, pp. 181-187.