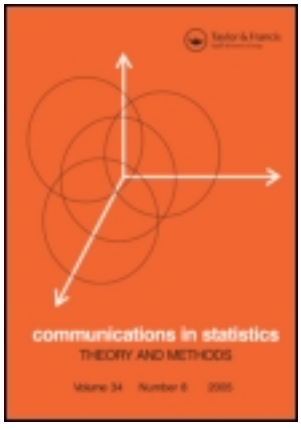
A Dendrite Method for Cluster Analysis

Article · January 1974

DOI: 10.1080/03610927408827101

2 authors, including:

Tadeusz Caliński
Poznań University of Life Sciences
59 PUBLICATIONS **2,534** CITATIONS



Communications in Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta19

A dendrite method for cluster analysis

T. Caliński [a] & J Harabasz [a]
[a] Academy of Agriculture , Poznań, Poland
Published online: 27 Jun 1974.

COMMUNICATIONS IN STATISTICS, 3(1), 1-27 (1974)

A DENDRITE METHOD FOR CLUSTER ANALYSIS

T. Caliński and J. Harabasz

Academy of Agriculture, Poznań, Poland

Key Words & Phrases: numerical taxonomy; cluster analysis; minimum variance (WGSS) criterion for optimal grouping; approximate grouping procedure; shortest dendrite = minimum spanning tree; variance ratio criterion for best number of groups.

ABSTRACT

A method for identifying clusters of points in a multi-dimensional Euclidean space is described and its application to taxonomy considered. It reconciles, in a sense, two different approaches to the investigation of the spatial relationships between the points, viz., the agglomerative and the divisive methods. A graph, the shortest dendrite of Florek et al. (1951a), is constructed on a nearest neighbour basis and then divided into clusters by applying the criterion of minimum within-cluster sum of squares. This procedure ensures an effective reduction of the number of possible splits. The method may be applied to a dichotomous division, but is perfectly suitable also for a global division into any number of clusters. An informal indicator of the "best number" of clusters is suggested. It is a "variance ratio criterion" giving some insight into the structure of the points. The method is

illustrated by three examples, one of which is original. The results obtained by the dendrite method are compared with those obtained by using the agglomerative method of Ward (1963) and the divisive method of Edwards and Cavalli-Sforza (1965).

1. INTRODUCTION

Various methods have been proposed for identifying groups of points in multidimensional spaces. The demand for such methods comes specially from systematists engaged in classificatory or taxonomical problems, in which each of the multivariate individuals under study may be considered as a point in a multidimensional space with an assigned distance measure. Thus classification of individuals consists in grouping of points. These groups are often called clusters, though no satisfactory definition of this concept exists. Its intuitive meaning is "that points within a cluster are close together, while the clusters themselves are far apart" (Rao, 1964, p. 351). The lack of a precise definition of such clusters as well as the computational difficulties in finding absolute optimal groupings give rise to many different approaches to cluster analysis and so to the application of various techniques.

Two methods of cluster analysis may differ in the choice of a measure of homogeneity within clusters and of heterogeneity between clusters, or in the procedure of applying this measure in grouping points into clusters, or in both. A functional relation chosen as a measure of the within-cluster homogeneity (or the between-cluster heterogeneity) usually re-

flects the relative desirability of grouping and depends on the nature of the problem. This objective function,as it is sometimes called (cf. Ward, 1963), does not, however, determine a method of cluster analysis. This usually depends also on the algorithm by which the clusters are constructed to optimize the objective function.Since cluster analysis is often applied in large-scale studies, the algorithm must not only be consistent with the criterion reflected by the objective function but also feasible in practical application to extensive data. Often a precise optimal solution for a well defined objective function is not possible, for the amount of computation involved becomes enormously large even with a moderate number of individuals. In such circumstances a non--exhaustive approximate procedure allowing for a reduction in computations must be devised. This may be done in different ways and so various techniques are suggested.

A familiar objective function applicable in cluster analysis is the within-group (cluster) sum of squares (WGSS). It seems natural to regard the optimal grouping of n points into k clusters as that for which WGSS is minimized. This criterion reflects a desire to find some minimum variance spherical clusters.

However attractive, the application of WGSS as an objective function demands the examination of all possible groupings of n points into k clusters and thus becomes impracticable even for small values of n. For example, the grouping of 10 points into 5 clusters requires 42 525 possibilities to be examined, and this number increases rather rapidly with the

rise in n (cf. Fortier and Solomon, 1966, section 1). Therefore, it is important to have a strategy that would reduce efficiently the number of computations. Among various proposals, two strategies seem to have gained particular interest among taxonomists. Although different in approach, they both employ sequential procedures and lead to hierarchical groupings. One of the strategies is the algorithm proposed by Ward (1963). Its idea is to agglomerate the points or the resulting clusters by reducing their number by one at each stage of a sequential fusion procedure, until all points are in one cluster. Given k clusters at a stage, k(k-1)/2 possibilities have to be examined for the reduction to k-1 clusters. A contrary algorithm has been suggested by Edwards and Cavalli-Sforza (1965). The essence of their method is the consecutive partition of a set of points into two subsets: first an initial set is divided into two clusters, then each of them is subdivided into two smaller clusters separately, and so on, until individual points are reached. For the division of n points into two clusters there are $2^{n-1} - 1$ possible partitions to be examined. A striking, though not unexpected, feature of the two methods is that operating with exactly the same minimum WGSS criterion they don't, in general, lead to the same hierarchical groupings. This in particular is the result when the points do not form well-separated clusters. The reason for the inconsistency is the obvious fact that any grouping at a stage of a sequential procedure is partly determined by the earlier stages. Moreover, the clustering obtained by a sequential method may, for the same

reason, differ considerably from the result obtainable by an exact global procedure (cf. Fortier and Solomon, 1966, p. 503).

In this paper another strategy for reducing the computing load is devised. It is based on the application, as an ancillary objective function, of the total length of trees spanning all points of the examined clusters. It may be shown that using as a criterion of grouping the minimum of this function, the same result is obtained by any of the described sequential procedures, i.e. by agglomeration or by division. (Differences may only emerge from the occurrence of a choice of several different tree edges of equal minimum length.) The property of a unique solution for the grouping based on the suggested tree function becomes evident when we recall that a tree spanning a set of points is "a connected graph that has no circuits " (cf., e.g., Ore, 1963, chapter 3). It follows from this definition that several trees may be connected in one tree and, vice versa, a tree may be disconnected into a number of separate trees. Furthermore, if points are connected into trees in such a way that the minimum of the total length of the tree edges is observed throughout the procedure, a shortest possible tree spanning all points results. Again, trees obtained from the shortest tree by consecutive removal of the longest edges will always ensure the minimum of the total length. A rigorous proof of this may be found in Florek et al. (1951a). More recently, properties of such trees have been discussed by Gower and Ross (1969). The former authors use the term dendrite instead of tree and are

concerned with the shortest dendrite. Gower and Ross (1969) call it the minimum spanning tree (MST). Both terms, as synonyms, will be used in this paper.

The shortest dendrite method has already been applied to many taxonomical problems, first by Florek et al. (1951b), and then by their followers (cf. the reviews given by Perkal, 1953, 1963). A cluster analysis based entirely on the shortest dendrite is known in Poland as "Taksonomia Wrocławska" (Wrocław Taxonomy). An extensive review of various applications of the MST is given by Gower and Ross (1969), who also describe the most common algorithms for finding the MST. Algol 60 algorithms for computing and printing the MST have been written by Ross (1969a, b).

In the approach to cluster analysis presented in this paper (as well as in an earlier paper by Caliński, 1969) the construction of the shortest dendrite is merely a starting point for a minimum variance partition. It reduces the enormous number of all possible partitions of a set of points to those only which are obtainable from a split of the shortest dendrite. Since the shortest dendrite ensures that each point is connected with its nearest neighbour (i.e. with that to which it has the smallest distance), the clustering of points from the same branch of the shortest dendrite will usually contribute to the WGSS less than the clustering of points from different branches. Hence, the limitation of possible groupings to the optimal splits of the shortest dendrite eliminates in advance most of the "poor" groupings, i.e. those with higher values of WGSS. It may happen that, together

with the poor groupings, also the absolute optimal grouping (with the minimum WGSS) will be eliminated. This is likely, however, only when the points are poorly separated into clusters.But even in this case the sacrifice of the absolute best grouping may be worth the considerable saving in computation. In fact no global procedures for cluster analysis that could ensure the finding of a precise optimal grouping exist and only methods that give a nearly optimal solution are possible (cf. Bolshev, 1969).

2. THE METHOD

Suppose there are n individuals (or samples from n populations) with observations on the same v variates for each individual. We may imagine them as being represented by n points in a v-dimensional Euclidean space, $P_1, \ldots, P_n$. The character of the variates is not essential for this representation, provided a measure of the distances between the individuals is well defined. It permits the computation of an n x n distance matrix, i.e. the $\underline{Q}$ matrix of Gower (1966), which is essential for the starting point of our method. Though the criteria we use are based on certain sums of squares, it is not necessary to calculate a v x v dispersion matrix of the points, i.e. the $\underline{R}$ matrix of Gower (1966).

If we denote the original v x n data matrix by $\underline{X}$, with rows given by the observed variates and with columns given by the individuals, we can write $\underline{X} = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n)$, where the column $\underline{x}_i$ is a vector of the v co-ordinates of the point P_i. If we refer the co-ordinates to orthogonal axes of an ordinary Euclidean space then the distance d_{ij} between P_i and

P_j will be properly defined by the function

$$d_{ij}^2 = (\underline{x}_i - \underline{x}_j)'(\underline{x}_i - \underline{x}_j), \quad i,j = 1,2,\ldots, n.$$

A similar formula applies to the distance between a point and the centroid of the n points. In the approach to cluster analysis which we follow, the dispersion of a group of n points is measured by the sum of the squared distances of the points from their centroid (cf. Gower, 1967). This sum is equal to the trace of the matrix $\underline{R}$, but may be obtained from the pairwise distances d_{ij} by applying the formula

$$\text{Trace } \underline{R} = n^{-1} (d_{12}^2 + d_{13}^2 + \ldots + d_{n-1,n}^2). \qquad (1)$$

This is a useful formula avoiding the computation of the $\underline{R}$ matrix. The same formula holds when the co-ordinates of points are referred to oblique axes of a Euclidean space with an appropriate inner product, and thus with an appropriate distance function. This includes the case of points representing samples rather than individuals with distances between their means defined by Mahalanobis generalized distance (D^2). There are good reasons to extend the measure of dispersion given by the right side of (1) to other distance functions, even if they are not defined in terms of the inner products of Euclidean spaces.

As stated, we start in any case with the distance matrix $\underline{Q}$ and construct the shortest dendrite or MST (cf. Florek et al.,1951a, or Gower and Ross, 1969). This is then partitioned by removing some of its edges: k - 1 if we want to divide the n points into k groups. The sum of squares criterion is calculated for each of the $\binom{n-1}{k-1}$ possible splits. If

we examine a split leading to a division of the n points into k groups of n_1, n_2, ..., n_k points ($n_1 + n_2 + \ldots + n_k = n$), then the (pooled) WGSS is calculated by applying the right hand side of (1) to each of the clusters separately and then summing the results. For ordinary Euclidean space the same result would be obtained by the analysis-of-variance partition of the matrix $\underline{R}$ into parts corresponding to the dispersion between and within the clusters of points, $\underline{R} = \underline{B} + \underline{W}$, and then taking the trace of $\underline{W}$ (cf. Friedman and Rubin, 1967, p.1163). We may then write

$$\text{WGSS} = \text{Trace } \underline{W} = \text{Trace } \underline{R}_1 + \text{Trace } \underline{R}_2 + \ldots + \text{Trace } \underline{R}_k,$$

where

$$\text{Trace } \underline{R}_g = n_g^{-1}(d^2_{12(g)} + d^2_{13(g)} + \ldots + d^2_{n_g-1,n_g(g)}),$$

with $d_{ij(g)}$ denoting the distance between points P_i and P_j in the g-th cluster (g = 1, 2, ..., k). Since one could extend the proposed method to cases where the points are not supposed to be in an ordinary Euclidean space and the dispersion matrices $\underline{R}$, $\underline{B}$ and $\underline{W}$ ($= \underline{R}_1 + \ldots + \underline{R}_k$) might have little meaning, we shall use the traditional notation of WGSS for Trace $\underline{W}$, BGSS (between-group sum of squares) for Trace $\underline{B}$ and TSS (total sum of squares) for Trace $\underline{R}$.

Consistently with the principle of the minimum variance criterion we decide on that partition of the shortest dendrite into k clusters for which WGSS is a minimum. But unlike Edwards and Cavalli-Sforza (1965) we search only among the $\binom{n-1}{k-1}$ partitions, instead of the much larger total of possibilities (as seen in Fortier and Solomon, 1966).

If k, the number of clusters, is not known, we proceed as follows: first we take k=2, then k=3, and so on. At each stage we find "the best sum of squares split" of the dendrite, for which we calculate not only the (minimum) WGSS, but also the (maximum) BGSS and the variance ratio criterion

$$\text{VRC} = \frac{\text{BGSS}}{k-1} \Big/ \frac{\text{WGSS}}{n-k} . \qquad (2)$$

We suggest the application of (2) as an informal indicator for the "best number" of groups. It is evident that this criterion is analogous to the F-statistic in univariate analysis. In fact it has already been used by Edwards and Cavalli-Sforza (1965, p. 374) as an F-test in a multivariate cluster analysis.

Though there is no satisfactory probabilistic theory to justify the use of VRC (2), the criterion has some desirable mathematical features that are encouraging. If $\bar{d}^2$ denotes the general mean of all $n(n-1)/2$ squared distances d_{ij}^2, and $\bar{d}_g^2$ that of the $n_g(n_g-1)/2$ squared distances within the g-th group ($g = 1, 2, \ldots, k$), then, from (1),

$$\text{TSS} = \frac{1}{2}(n-1)\bar{d}^2,$$

$$\text{WGSS} = \frac{1}{2}((n_1-1)\bar{d}_1^2 + (n_2-1)\bar{d}_2^2 + \ldots + (n_k-1)\bar{d}_k^2)$$

and

$$\text{BGSS} = \frac{1}{2}((k-1)\bar{d}^2 + (n-k)\,A_k),$$

where

$$A_k = \frac{1}{n-k}((n_1-1)(\bar{d}^2-\bar{d}_1^2) + (n_2-1)(\bar{d}^2-\bar{d}_2^2) + \ldots + (n_k-1)(\bar{d}^2-\bar{d}_k^2))$$

is a weighted mean of the differences between the general and the within-group mean squared distances. Now we may write

$$\text{VRC} = \frac{\text{BGSS}}{k-1} \Big/ \frac{\text{WGSS}}{n-k} = (\bar{d}^2 + \frac{n-k}{k-1} A_k)/(\bar{d}^2 - A_k).$$

It is evident that in the special case of equal distances between all pairs of points A_k becomes zero and VRC is one. Otherwise the minimum WGSS criterion maximizes A_k for a given k. As an average, the function A_k may also be used to compare partitions obtained for different numbers of groups: the difference $A_k - A_{k-1}$ will indicate an average gain in the within-group compactness resulting from the change from k - 1 to k groups. Hence, the behaviour of A_k as a function of k may be sensitive to the existence of groups. To see this in connection with the VRC it is instructive to write

$$\frac{BGSS}{k-1} \Big/ \frac{WGSS}{n-k} = \left(1 + \frac{n-k}{k-1} a_k\right) \Big/ (1 - a_k), \qquad (3)$$

where

$$a_k = A_k / \bar{d}^2.$$

Since the minimum WGSS (and so the maximum BGSS) is used in (3), we have a_k between 0 and 1, with $a_k = 0$ for equal distances between all pairs of points and with $a_k = 1$ for an "ideal" clustering, i.e. for no variation within groups. (If all points are different this is not obtained earlier than at the final stage of k = n.) If the points are uniformly distributed in space, a_k will increase slowly and more or less steadily with the rising value of k. And from (3), the VRC will tend to decrease when k increases and a_k is constant, this being more or less counterbalanced by the increase in a_k. Anyway, a uniform distribution of points in space will be usually reflected by a smooth run of values of the VRC. On the other hand, if the points are grouped into k_0 natural clusters, with a small within-cluster variation, the change from $k_0 - 1$ to k_0 will cause a considerable increase in a_k

and so a rapid rise of the VRC, possibly forming a hump. More precisely, the increase in the number of groups from $k_0 - 1$ to k_0 will cause an increase of the VRC if a_{k_0}/a_{k_0-1} exceeds the ratio $(k_0 - 1)/(a_{k_0-1} + k_0 - 2)$ which is never smaller than one.

It follows from the discussion above that the computation of VRC for k = 2, 3, ... may be helpful in deciding on the "best number" of groups. We suggest choosing that number k for which the VRC has an absolute or local maximum, or at least has a comparatively rapid increase. If there are several such local maxima, it will be most economical to choose the smallest of the related values of k. This in fact means that the computation can be stopped when the first local maximum is reached. The process may then be repeated for each of the resulting groups separately, and so on. This further suggests that the dichotomous grouping of Edwards and Cavalli-Sforza (1965) is advisable when the first values of VRC form a monotonic decreasing sequence. Also it seems that when the values of VRC are increasing monotonically throughout the range of k, then no reasonably better partition of the points exists than that into individuals.

3. THE COMPUTER PROGRAMS

Several computer programs are available for the dendrite method. Our own programs are written in Most I for Odra 1013, in Mat IV for Mińsk 22 and in Algol for Odra 1204. Fortran programs have been written by Wishart (1970) - a Fortran II program for IBM 1620 and a Fortran IV program for IBM 360. They are included in the CLUSTAN IA suite of Fortran programs

for cluster analysis and other multivariate procedures, distributed by the St. Andrews University Computer Laboratory, Scotland.

The programs compute and print the shortest dendrite and then divide it into 2, 3, ..., n-1 clusters on the minimum WGSS basis. A minimum and a maximum for cluster numbers that are of interest may be specified, to limit the computations. This option is important, since for n greater than 20 the execution of all optimal division from k = 2 to k = n - 1 may require considerable computing time. Several grouping criteria are computed and printed, including the suggested VRC which helps to decide on the "best number" of clusters within the specified range.

The programs are restricted to n not greater than 140.

4. EXAMPLES

The main purpose of presenting the following examples is to compare the dendrite method with the sequential methods of Ward (1963) and of Edwards and Cavalli-Sforza (1965). It is also hoped that the examples will make more explicit the idea of linking points into dendrites for a cluster analysis.

4.1. Bacteriological Data

The data of this example consist of a number of scores observed for six species of bacteria and are taken from Edwards and Cavalli-Sforza (1965). Table I gives the $\binom{6}{2}$ squared distances between the six species in an ordinary Euclidean space. The shortest dendrite constructed on the basis of these distances is shown in Figure 1. It may be obtained in the following way: We start by choosing the shortest dis-

tance between the points (species),i.e. the distance between A and B. It forms the first edge of the dendrite, A-B. Then the shortest edge which connects to A-B is added, i.e. B-D. Now the dendrite consists of two edges, A-B-D. It is then extended by adding the shortest of the remaining edges which connects to at least one edge of the present dendrite without forming a circuit, i.e. D-C. The dendrite is now of the form

TABLE I

The Half-Matrix of Squared Distances for the Bacteriological Data

A	B	C	D	E	F	Points
	5	11	11	14	14	A
		10	6	13	15	B
			6	17	21	C
				13	15	D
					6	E
						F

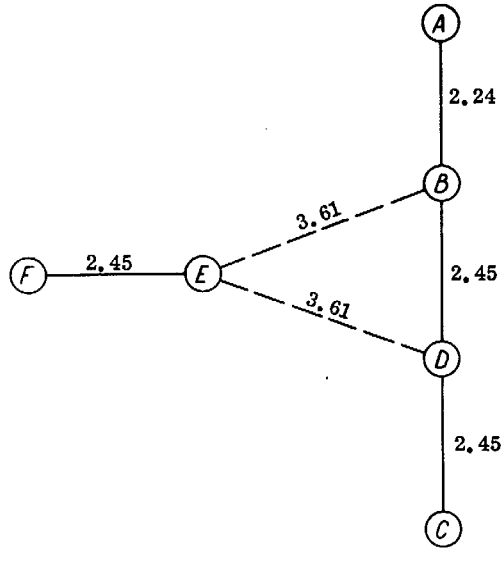


FIG. 1

The shortest dendrite for the bacteriological data.

A-B-D-C. Following the same rule of extension, the edge B-E or D-E is to be added, since both are of equal length and shorter than any other to be considered. These two possibilities are indicated in Figure 1 by two alternative broken lines. Therefore, the graph drawn in Figure 1 should be interpreted as giving two alternative shortest dendrites with either the edge B-E or D-E. The dendrite is completed by adding E-F, the shortest edge which connects the remaining point F. The lengths of the edges, that refer to the distances between the connected points, are also given in Figure 1.

The resulting shortest dendrite (MST) will now be split into most compact groups of points. This may be done in the same way as in Edwards and Cavalli-Sforza (1965), except that only those partitions are examined which emerge after removing one edge from the dendrite at each stage of the subdivision. This leads to exactly the same splits as those obtained by Edwards and Cavalli-Sforza (1965), who examined all possible ways of subdividing the points into two clusters at a given stage of the sequential procedure. The partition into clusters of ABCD and EF results from the removal of the edge E-B or E-D (depending on which one is included in the dendrite), after examining only 5 out of the 31 possible splits. Further subdivision into clusters of AB and CD is achieved by trying only 3 out of the 7 possible splits.

But the clustering needs not to be restricted to the dichotomous subdivision of the set of points. If so desired, we may split the dendrite into a "given number" of groups or decide on the "best number" of groups by examining the behav-

iour of the VRC (as described in section 2).

The division into k = 2 groups has already been discussed, the best split is ABCD: EF. The division into k = 3 groups results from removing 2 edges from the dendrite. Applying the minimum WGSS criterion, AB: CD: EF is obtained as the best split. To receive the partition into k = 4 groups we remove 3 edges, obtaining in this case two possible best splits, AB: C: D: EF or AB: CD: E: F. Finally, for k = 5 we obtain as the best split AB: C: D: E: F. It has been found that when examining all possible splits into k = 2, 3, 4 and 5 groups with the minimum WGSS criterion the results are exactly the same as those just presented. Table II summarizes the grouping criteria of the resulting splits. It reveals some hierarchical structure of the data. This conclusion is drawn from the VRC given in the last row. It suggests that the best split is obtained with two groups, which in this case is directly evident in the distance table (Table I): E and F are far apart from the rest of the points. This also explains the agreement between the results of the dendrite method and those obtained by Edwards and Cavalli-Sforza (1965). The full agreement of the results of all three compared methods and the exact global method may be explained by the apparent good separation of the hierarchically-built clusters.

4.2. Anthropometric Data

In this example we reexamine the anthropometric data originally analyzed by Rao (1952) and used as an example also by Edwards and Cavalli-Sforza (1965). Here the distances between points which represent sample means of nine anthropo-

TABLE II

Criteria for the Cluster Analysis of the Bacteriological Data

Number o groups	2	3	4	5
Number of possible splits:				
total	31	90	65	15
in dendrite method	5	10	10	5
Max BGSS[a]	14.25	21.0	24.0	27.0
Min WGSS[a]	15.25	8.5	5.5	2.5
$VRC = \frac{n-k}{k-1} \cdot \frac{\text{Max BGSS}}{\text{Min WGSS}}$	3.74	3.70	2.91	2.70

[a] The same for Ward method, Edwards and Cavalli-Sforza method, dendrite method and the exact global method.

metric characters for twelve Indian castes and tribes are defined by Mahalanobis generalized distance. They are given in Table IV of Edwards and Cavalli-Sforza (1965). The shortest dendrite based on the distances is given in Figure 2, in two different versions. The dendrite (a) on the left has been drawn in the usual way, with the length of its edges proportional to Mahalanobis D. The dendrite (b) on the right has been drawn in a plane with the first two principal components as coordinate axes (as found by Rao, 1952, chapter 9c). This presentation slightly distorts the lengths of the edges. Real values of D are given in both of the insomorphic dendrites.

Without any a priori decision on the number of clusters, we have performed the whole sequence of calculations described in section 2. Some of the grouping criteria are given in Table III. One interesting point is the reduction of the number of partitions to be examined by the minimum WGSS cri-

FIG. 2

The shortest dendrite for the anthropometric data drawn (a) in the usual way and (b) in the chart of the first two principal components (canonical variates).

terion. Another striking point is the pattern of the sequence of values obtained for the VRC: it has a greater value for k = 5 than for k = 4 or 6. This is the only disturbance in the monotonic increase of this criterion for rising values of k. Following the suggestion given at the end of section 2, we should then take k = 5 as the "best number" of groups of points into which to split the shortest dendrite. The WGSS splitting of the shortest dendrite of the anthropometric data leads to the following five clusters: (I) Ahir, Kurmi, Other Artisan, Kahar; (II) Chattri, Muslim; (III) Dom, Bhil; (IV) Basti Brahmin, Other Brahmin; (V) Bhatu, Habru. This clustering is exactly the same as that obtained by Rao (1952), who arrived at it largely by intuition supported by an average distance criterion. The same result has also been obtained by the method of Ward (1963).

A different grouping of these data was found by Edwards and Cavalli-Sforza (1965), who used the minimum WGSS criterion in a dichotomous subdivision of the set of points. But since the values of the VRC, as given in Table III, form an almost consistently increasing sequence, there is no reason to assume a hierarchical structure of the points. Therefore, a method which sequentially divides the set into two groups at each stage of the procedure is here unjustified. As the result of such inappropriate grouping the Ahir have been clustered with the Brahmin, though the former are evidently nearer to the Kurmi. Edwards and Cavalli-Sforza (1965) are aware of this difficulty but consider unfeasible the examination of all possible splits into more than two groups. The

TABLE III

Criteria for the Cluster Analysis of the Anthropometric Data

Number of groups	2	3	4	5	6	7	8	9	10	11
Number of possible splits:										
total	2047	86526	611501	1379400	1323652	627396	159027	22275	1705	66
in dendrite method	11	55	165	330	462	462	330	165	55	11
Min WGSS:										
in Ward method	187.12	132.12	83.33	46.25	33.75	22.75	12.17	7.50	3.50	1.00
in Edwards and Cavalli-Sforza method	180.69	126.60	83.70	48.50	36.00	24.50	13.50	7.50	3.50	1.00
in dendrite method	180.69	126.60	80.63	46.25	33.75	22.75	12.17	7.50	3.50	1.00
VRC in dendrite method	3.87	4.41	5.62	7.74	7.71	8.35	11.20	12.16	15.69	24.97

considerable reduction of the number of possible ways gained by the dendrite method removes this difficulty and avoids dichotomous clustering not justified by data. The full agreement of the proposed method with the result of Rao (1952) supports the suggested VRC of the "best number" of groups. Furthermore, judging from the values of the minimum WGSS criterion given in Table III for all three compared methods, the dendrite method has appeared superior to the method of Ward (1963) for k = 2, 3 and 4, and superior to that of Edwards and Cavalli-Sforza (1965) for k = 4, 5, 6, 7 and 8. In no case has the dendrite method been inferior to any of the sequential methods.

4.3. Plant Breeding-Data

The data analyzed in this subsection (collected by Dr Z. Kłoczowski of the Institute of Plant Breeding and Acclimatization, Poznań) consist of 4 measurements on 30 flowers from each of 7 strains of sunflower. Table IV gives the Mahalanobis distances (D^2) between the strains computed for all measurements. At the bottom of the table the smallest significant squared distances at the 5% and 1% levels are given. They are computed from Hotelling's T^2-distribution. The shortest dendrite based on Table IV is given in Figure 3. The numbers of possible splits together with the calculated criteria are presented in Table V. The VRC suggests a split into r = 5 groups. The splitting of the dendrite that minimizes WGSS

TABLE IV

The Half-Matrix of Squared Mahalnobis Distances for the Plant Breeding Data

A	B	C	D	E	F	G	Points
	1.42	0.36	1.93	1.23	3.57	5.52	A
		1.25	4.49	1.10	2.40	4.04	B
			1.23	1.12	2.86	4.93	C
				3.96	4.45	6.66	D
					1.90	3.13	E
						0.31	F
							G

$D^2_{0.05} = 0.66$ $D^2_{0.01} = 0.92$

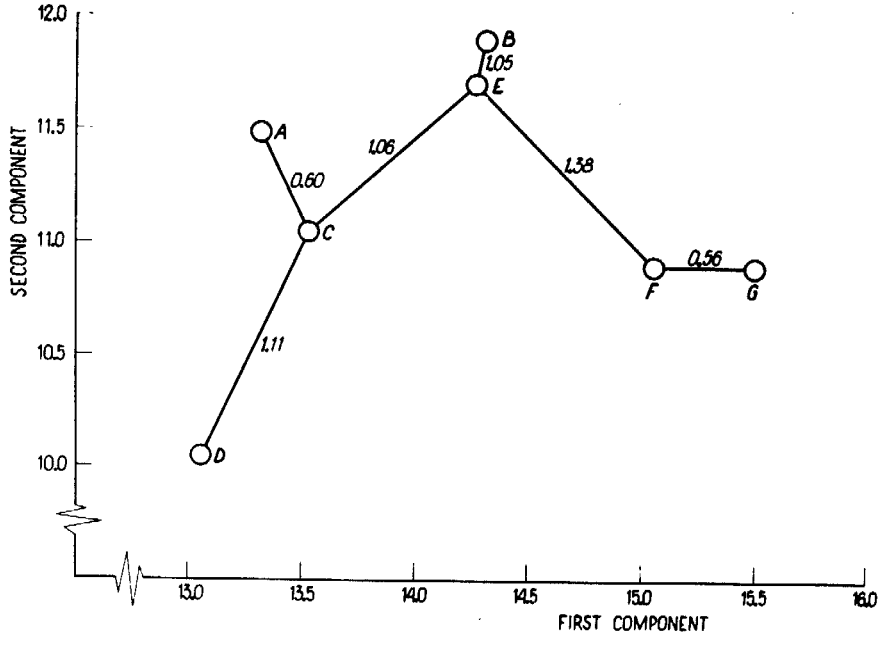


FIG. 3

The shortest dendrite for the plant breeding data.

TABLE V

Criteria for the Cluster Analysis of the Plant Breeding Data

Number of groups	2	3	4	5	6
Number of possible splits:					
total	63	301	350	140	21
in dendrite method	6	15	20	15	6
Max BGSS[a]	4.49	6.49	7.38	7.93	8.11
Min WGSS[a]	3.77	1.78	0.89	0.34	0.16
VRC = $\frac{n-k}{k-1}\cdot\frac{BGSS}{WGSS}$	5.95	7.31	8.34	11.84	10.47

[a]The same for Ward method, Edwards and Cavalli-Sforza method, dendrite method and the exact global method.

(= 0.34) is D: AC: B: E: FG. We notice that the two pairs of strains that have not been separated in the split are the only ones that are not significantly different (cf. Table IV). It is also evident from TableIV that no other division of the seven strains into five clusters could give a smaller WGSS than the one based on the shortest dendrite. In fact it has been found that all three compared methods give here the same results, which are in complete agreement with the results obtained from examining all possible splits.

Finally, it may be interesting to compare the sequences of values of the VRC (2) calculated for the three examples under consideration. They are given in the charts of Figure 4. The criterion for the bacteriological data (a) results in a decreasing sequence thus suggesting a possible hierarchical structure of the points. The sequence for anthropometric data (b) is increasing, except for a hump at k = 5. This suggests that the clusters are not well separated and if there

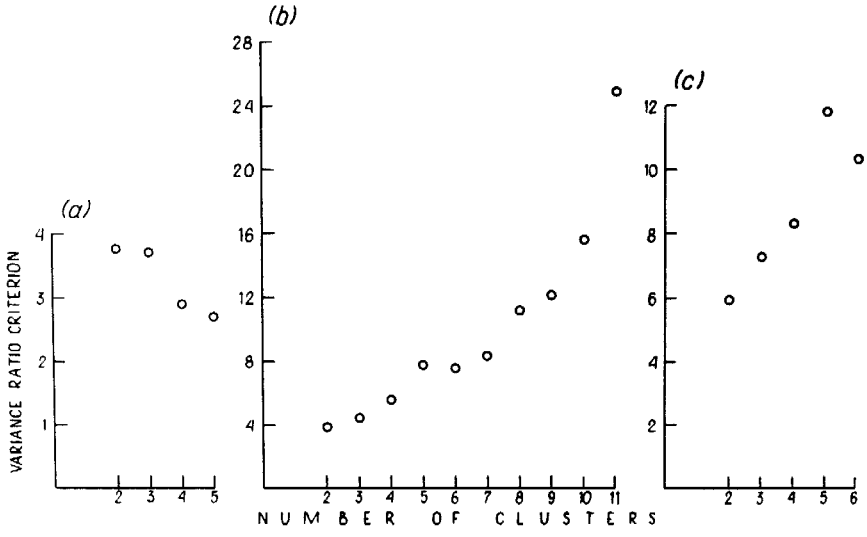


FIG. 4

The behaviour of the variance ratio criterion for the (a) bacteriological, (b) anthropometric and (c) plant breeding data.

are any clusters the most likely number of them is five. The behaviour of the criterion is particularly evident in the sunflower case (c), suggesting the clustering into 5 groups.

These examples do not exhaust all possible performances of the VRC, and we are aware that in some cases the decision on the "best number" of groups based on VRC may be vague. But the cases we have investigated show that this criterion gives some instructive insight into the structure of points. To make the application of the criterion more precise, further studies on the geometrical distributions of points in Euclidean spaces would be necessary.

5. CONCLUSION

Data consisting of v measurements on each of n objects (individuals or samples) may meaningfully be subjected to cluster analysis if a measure of pairwise distance between objects is well defined. The objects may then be thought of as points in a Euclidean space and the compactness of clusters of them may be measured by the sums of the squared distances of the points from the centroids of the clusters. The minimum of the WGSS becomes then an appropriate criterion for cluster analysis. The enormous number of possible ways of dividing the set of n points into k groups can effectively be reduced by constructing and splitting the shortest dendrite. This dendrite method appears to be a suitable and satisfactory approximate procedure. The "best number" of groups, k, can often be determined by the VRC, which also gives same insight into the spatial structure of the points, which need not necessarily be hierarchical. The proposed method does not impose any such structure on the resulting clusters. In this sense it is more widely applicable than many other methods of cluster analysis. Particularly, it seems to be superior to the sequential procedures of Ward (1963) and Edwards and Cavalli-Sforza (1965), or at least advisable as a method complementary to them. Performing and comparing all the three procedures it is always possible to decide on that grouping which gives the lowest WGSS.

ACKNOWLEDGMENTS

We wish to thank Dr. David Wishart for writing the Fortran programs and for including the dendrite method into his CLUSTAN IA package.

BIBLIOGRAPHY

Bolshev, L.N. (1969). Cluster analysis. Proc. 37th Session of the International Statistical Institute, Book 1, 411-25.

Caliński, T. (1969). On the application of cluster analysis to experimental results. Proc. 37th Session of the International Statistical Institute, Book 2, 108-10.

Edwards, A.W.F. and Cavalli-Sforza, L.L. (1965). A method for cluster analysis. Biometrics 21, 362-75.

Florek, K., Łukaszewicz, J., Perkal, J., Steinhaus, H. and Zubrzycki, S. (1951a). Sur la liaison et la division des points d'un ensemble fini. Colloquium Mathematicum 2, 282-5.

Florek, K., Łukaszewicz, J., Perkal, J., Steinhaus, H. and Zubrzycki S. (1951b). Taksonomia wrocławska. Przegląd Antropologiczny 17, 193-211.

Fortier, J.J. and Solomon, H. (1966). Clustering procedures. Multivariate Analysis (Ed. P.R. Krishnaiah), 493-506.Academic Press, New York.

Friedman, H.P. and Rubin, J. (1967). On some invariant criteria for grouping data. J. Amer. Statist. Ass. 62,1159-78.

Gower, J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53, 325-38.

Gower, J.C. (1967). A comparison of some methods of cluster analysis. Biometrics 23, 623-37.

Gower, J.C. and Ross, G.J.S. (1969). Minimum spanning trees and single linkage cluster analysis. Appl. Statist. 18, 54-64.

Ore, O. (1963). Graphs and their Uses. Random House, Inc., New York.

Perkal, J. (1953). Taksonomia wrocławska. Przegląd Antropologiczny 19, 82-105.

Perkal, J. (1963). Matematyka dla Przyrodników i Rolników. Państwowe Wydawnictwo Naukowe, Warszawa.

Rao, C.R. (1952). Advanced Statistical Methods in Biometric Research, John Wiley and Sons, Inc., New York.

Rao, C.R. (1964). The use and interpretation of principal component analysis in applied research. Sankhyā A 26, 329-58.

Ross, G.J.S. (1969a). Minimum spanning tree (Algorithm AS 13). Appl. Statist. 18, 103–4.

Ross, G.J.S. (1969b). Printing the minimum spanning tree (Algorithm AS 14). Appl. Statist. 18, 105–6.

Ward, Jr., J.H. (1963). Hierarchical grouping to optimize an objective function. J. Amer. Statist. Ass. 58, 236–44.

Wishart, D. (1970). 5 new Fortran IV programs for cluster analysis (CLUSTAN IA). Kansas Geologic Computer Contributions.

Received September 1972; retyped version received May 1973.

Recommended by N. L. Johnson, University of North Carolina at Chapel Hill.

Refereed by Lawrence S. Mayer, Virginia Polytechnic Institute and State University.