

H

ON THE ASSESSMENT AND COMPARISON OF CLASSIFICATIONS

A. D. Gordon

Department of Statistics, University of St. Andrews,
North Haugh, St. Andrews, Scotland.

SUMMARY

In classification studies, when different clustering criteria are used on the same data set the results are not usually identical; each clustering criterion appears to be effective at uncovering particular kinds of structure in data, and might impose unwarranted structure on a data set. A review is given of the reactions of various research workers to this problem. In particular, investigators have often undertaken a subjective synthesis of two or more classification schemes; where the schemes agree, the relationships depicted are less likely to be purely artifacts of the clustering criterion used. This paper describes more formal methods of comparing classification schemes to allow more detailed examination of areas in which they agree.

pages 149-160 in R. Tomassone (ed.)
Analyse de Données et Informatique,
I.N.R.I.A., Le Chesnay (1980).

[1980]

1. INTRODUCTION

- 1.1. Adaptive clustering schemes
- 1.2. Significance testing approach
- 1.3. Properties of clustering criteria
- 1.4. Comparative studies

2. COMPARING CLASSIFICATIONS

- 2.1. Comparing partitions and coverings
- 2.2. The intersection of trees
- 2.3. A pruning algorithm
- 2.4. Regrafting pruned branches

3. DISCUSSION

REFERENCES

1. INTRODUCTION

The explosion of interest in exploratory methods of multivariate data analysis in recent years has left scattered through the research literature a large number of classification algorithms [Cor71,Eve74,Har75]. New algorithms are still being proposed, but an increasing amount of effort is being devoted to the assessment and comparison of existing methods of classification; for example, R.R.Sokal [Sok77] opined :-

"Although clustering and ordination methods will undoubtedly be improved in the future, we seem to have at our disposal an adequate array of techniques to serve many purposes. I believe that the major effort in classificatory work in the next few years should be devoted to comparisons of different approaches and to tests of significance of classifications. Work in these fields has so far been quite unsatisfactory."

Little will be said in this paper on *ordination* methods, in which one seeks to represent the objects in the data set by a set of points in some low-dimensional space such that points which are close together correspond to objects which are similar to one another; we shall concentrate here on the assessment and comparison of *clustering* methods, *i.e.*,

- (1) methods which seek to impose on the set of objects either a partition into disjoint classes, or a covering of overlapping classes, such that objects within a class are similar to one another;
- (2) hierarchical methods, in which the similarities between objects are summarized in a nested set of partitions, which can be represented in a tree diagram, or dendrogram.

The words "classification scheme" will be used to refer collectively to the results of using such clustering methods.

It is worth noting that applied practitioners of the subject would appear to have reached a measure of consensus on which clustering algorithms to use; a study reported by R.K.Blashfield [Bla76] noted that nearly three-quarters of the published uses of cluster analysis in 1973 used one or more of three methods : single link, average link (UPGMA) [SoM58], and the minimum sum-of-squares criterion [War63].

The current interest in comparing clustering methods has been stimulated by the increasing awareness of the extent to which clustering criteria can misrepresent the structure in the data; for examples, see [Wis69,Eve74]. Each clustering criterion appears to be effective in detecting particular kinds of structure in data; for example, the sum-of-squares criterion tends to favour hyperspherical-shaped clusters, and the single link method will find clusters of any shape but requires that these clusters be distinct and not linked together by chains of intermediate objects. In short, each clustering criterion is pre-disposed to find clusters of a particular "shape", and may well distort the data to some extent. Various measures of distortion have been proposed [Cor71]; these can provide a warning against dubious conclusions, but a more detailed study would be needed to investigate what was causing a high measure of distortion.

Given that each clustering criterion may be well-suited to uncovering particular types of group structure, if we knew the "shape" of the clusters we might be able to select an appropriate clustering method (The fact that different clusters in the data could well have different "shapes" is an added complication). The circularity is immediately apparent : in general, we do not know the underlying structure in the data; the investigation is being undertaken precisely in order to find this out. The remainder of this section summarizes four directions from which this problem has been approached.

1.1. ADAPTIVE CLUSTERING SCHEMES

In adaptive clustering schemes [Roh70,Che72,DiG77], the expected cluster shape can be modified during the clustering procedure; this is usually achieved by basing the definition of within-cluster distance on the disposition of the current group members. To obtain reasonable results, the methods require each cluster to be composed of a fairly large number of objects. A limitation is that one must specify beforehand the class of cluster shapes considered permissible [Roh70]; future developments in this area could involve interactive or supervised programs, these being liable to the charge of greater dependence on the subjective (and difficult to specify precisely) choices of the user.

1.2. SIGNIFICANCE TESTING APPROACH

In this approach, researchers have posed the following question : "In the absence of any structure in the data, how would one expect a clustering criterion to behave?" The absence of structure can be regarded as a "null hypothesis" to be investigated. There have been difficulties in framing realistic null hypotheses which allow one to make much analytical progress. Some work [Lin73,BaH76] has assumed that the elements of the matrix of pairwise dissimilarities between objects are ranked in random order, and has derived the distribution of various statistics under this hypothesis. This assumption ignores the metric structure usually contained in the dissimilarities, but Ling argues [Lin73] that this model can be regarded as a limiting case and that if no significant clusters are indicated under this model then clusters are unlikely to be deemed significant under any model. One suspects that it will prove more difficult to make much analytical progress with more realistic null models, although recourse may be had to simulation studies [GoB75].

In carrying out a test of significance, one also has to frame an alternative hypothesis, and this has caused difficulties in this area. Clearly, an appropriate alternative hypothesis would be to assert that there is *some* structure in the data but this could take many forms, and tests to detect particular kinds of structure [EnH69,Har75,Boc77,Har77,Sne77a,Sne77b] need not be efficient at detecting other kinds of structure. It may be that the notion of "fuzzy clustering" [Bez74,Boc79] will prove useful here, in allowing one to test whether an observed "fuzzy" solution may reasonably be regarded as corresponding to a "hard" partition; alternatively, some data sets may be more formally postulated as arising from a multivariate mixture model [Day69,Wol70], although the multivariate normal distribution would seem to be inappropriate for many sets of data encountered in classification.

These more formal methods can be contrasted with the approach of many of the earlier workers in classification, who would have described their work as "exploratory data analysis" [Tuk77], or as the *generation* of hypotheses about the structure in the data [WiD65,Jar70], hypotheses which might then be tested *on new data*. The rationale is that data may well have many different properties; tests of hypotheses concentrate on a limited number of properties to the exclusion of others which might be of interest, so it might be wise not to resort to hypothesis testing too early in the analysis. This reasoning would lead to a two-stage process of analysis : "exploratory data analysis" followed by "confirmatory data analysis" [Tuk77] on different data. Few, if any, classification studies have proceeded to this second stage. In part, this has been due to the difficulty associated with collecting new data : sometimes this is not possible, for example when the set of objects is a collection of artifacts found at an archaeological site, and division of the original data into two subsets has not seemed appropriate; in part, it has been caused by the difficulty of framing an appropriate model and testing it on new data. The attitude adopted will depend on whether it is possible and useful to regard the data as a sample from some relevant "population", or whether one is only interested in the set of objects under study.

1.3. PROPERTIES OF CLUSTERING CRITERIA

Some of the work in the previous subsection can be described in terms of seeking to investigate and control errors of the first kind, *i.e.* situations where a criterion wrongly asserts that there is structure in the data. The work in this subsection is concerned with errors of the second kind: if there is structure in the data, we want the clustering criterion to detect it. There have been simulation studies to investigate the performance of several clustering criteria in detecting particular types of structure in data [BaH75, Dom75, KuF75]. This kind of approach can be of use in eliminating "almost always bad" methods, but it can provide only a limited amount of information. Even if one knew the most appropriate clustering method for every conceivable type of data set which might arise, the problem is that in general the precise form of the data is not known before the analysis begins.

In some situations, however, one has a certain amount of information on the expected properties of the data or the method of analysis. N.Jardine and R.Sibson [JaS71] show that if one requires the classification method to satisfy a specified list of axioms, then the *only* appropriate method of analysis is the single link method. An alternative approach is the admissibility approach of L.Fisher and J.W. Van Ness [FVN71, VNe73]: they have constructed a list of properties which one might expect the data to possess in some situations, and have investigated whether or not the groups obtained from various clustering methods would invariably possess these properties. The rationale is that although one may not be able to specify very precisely what properties one would expect of the data, one might have some relevant information from the manner in which the data arose. One could then use Fisher and Van Ness' admissibility table to select one or more *admissible* clustering criteria, regarding the other criteria as inappropriate for the particular data set being analysed.

1.4. COMPARATIVE STUDIES

The approaches considered thus far assume that one has some information about the data set, but often this is not available. The final approach accepts the fact that different criteria are predisposed to find particular kinds of structure, and so analyses a given set of data by several different methods. (In such studies, it is common to employ both clustering and ordination methods of analysis, but we will concentrate here on the comparison of the results from different clustering methods). If the classifications obtained using two or more different clustering criteria do not differ markedly, then one has more confidence in the reality of any group structure indicated; the results are less likely to be purely artifacts of a clustering method. Subjective assessments of several classifications have been commonly made [Min65, WWL66, GWW68, GrA69, Sch70, SnS73]. Other situations in which it might be relevant to compare two or more classifications - and which are investigated by most of the above-mentioned authors - are comparisons based on different measures of dissimilarity (possibly based on different sets of variables) relating to the same set of objects, and the comparison of a numerical classification with one based on external criteria. The remainder of this paper describes some more formal ways of synthesising the results from different classifications of the same data.

2. COMPARING CLASSIFICATIONS

The methods to be described in this section will be illustrated on a set of 16 objects [GPC66], where an object corresponds to a collection of botanical quadrats from Gotland which are dominated by a single species. Figs. 1 and 2 summarise the results of analysing these data using the single link and complete link methods. Sectioning the two tree diagrams at the places indicated by the dashed lines

yields the partitions into the four single link groups

S_1 (12), S_2 (1,5-7,15), S_3 (2,3,8-10,13,14,16), S_4 (4,11),
and the five complete link groups

C_1 (12), C_2 (1,5-7,15), C_3 (2,3,10,13), C_4 (8,16), C_5 (4,9,11,14).

Various statistics have been used to compare two classification schemes [GoK54, SoR62,Ran71], but these give a measure of the overall agreement between two classifications without directly indicating which objects are contributing towards this agreement. Some of these statistics can be interpreted in terms of fit and error distributions associated with the two classifications [Fra76]. The remainder of this section describes an approach aimed at identifying the "intersection" or "area of agreement" between two classification schemes : partitions and coverings are considered in subsection 2.1, and dendrograms are considered in subsections 2.2 to 2.4.

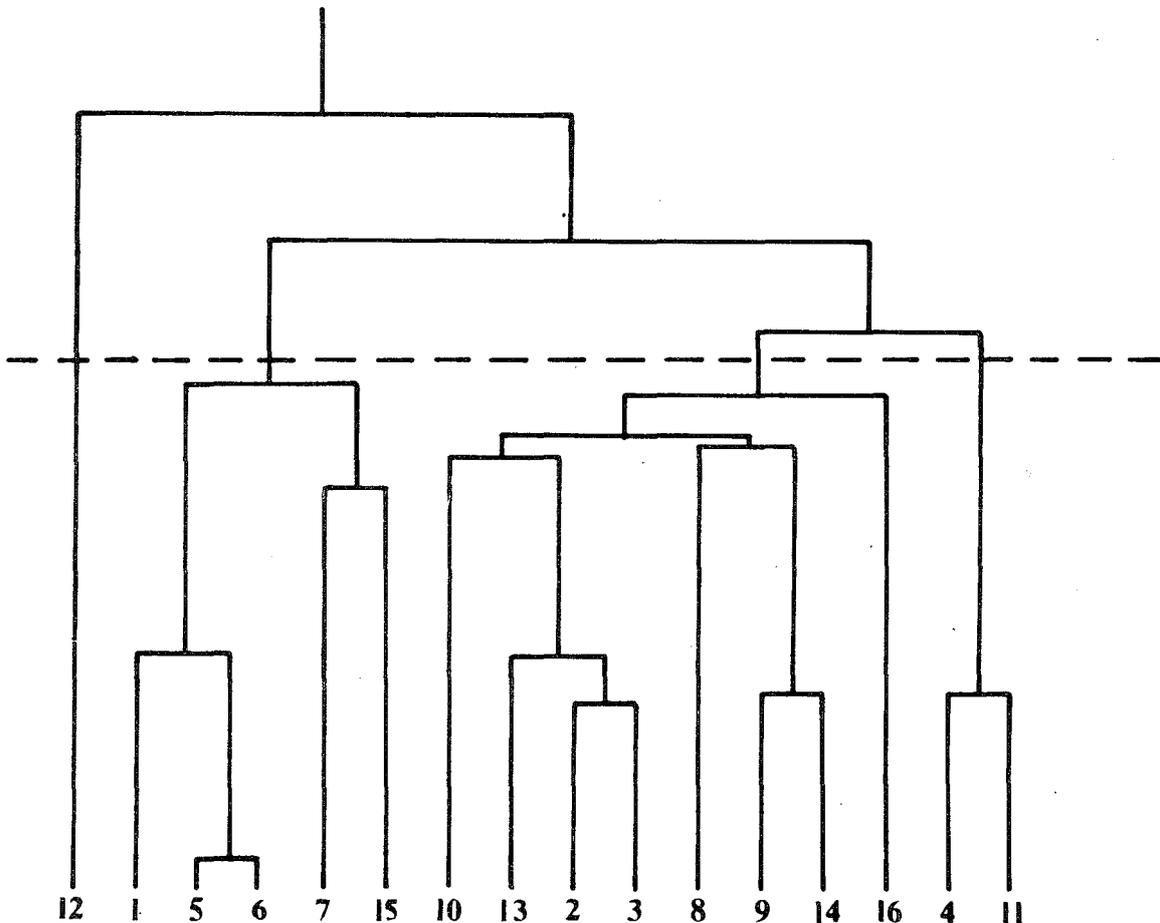


Fig. 1. Single link dendrogram of Gotland data.

2.1. COMPARING PARTITIONS AND COVERINGS

Table 1 summarises the number of objects occurring jointly in the single link and complete link groups $\{S_i, C_j \ (1 \leq i, j \leq 5)\}$ defined above; the empty group S_5 is introduced so as to equalise the number of groups in each partition. Rand's coefficient [Ran71] for these partitions takes the value 0.8, indicating a fair measure of agreement; perfect agreement would be given by a value of 1. In order to determine what is causing this agreement, match the groups in pairs, with one group in a pair from each partition, so as to maximise the number of objects in common in the set of matched pairs of groups. This is just the assignment problem in operational

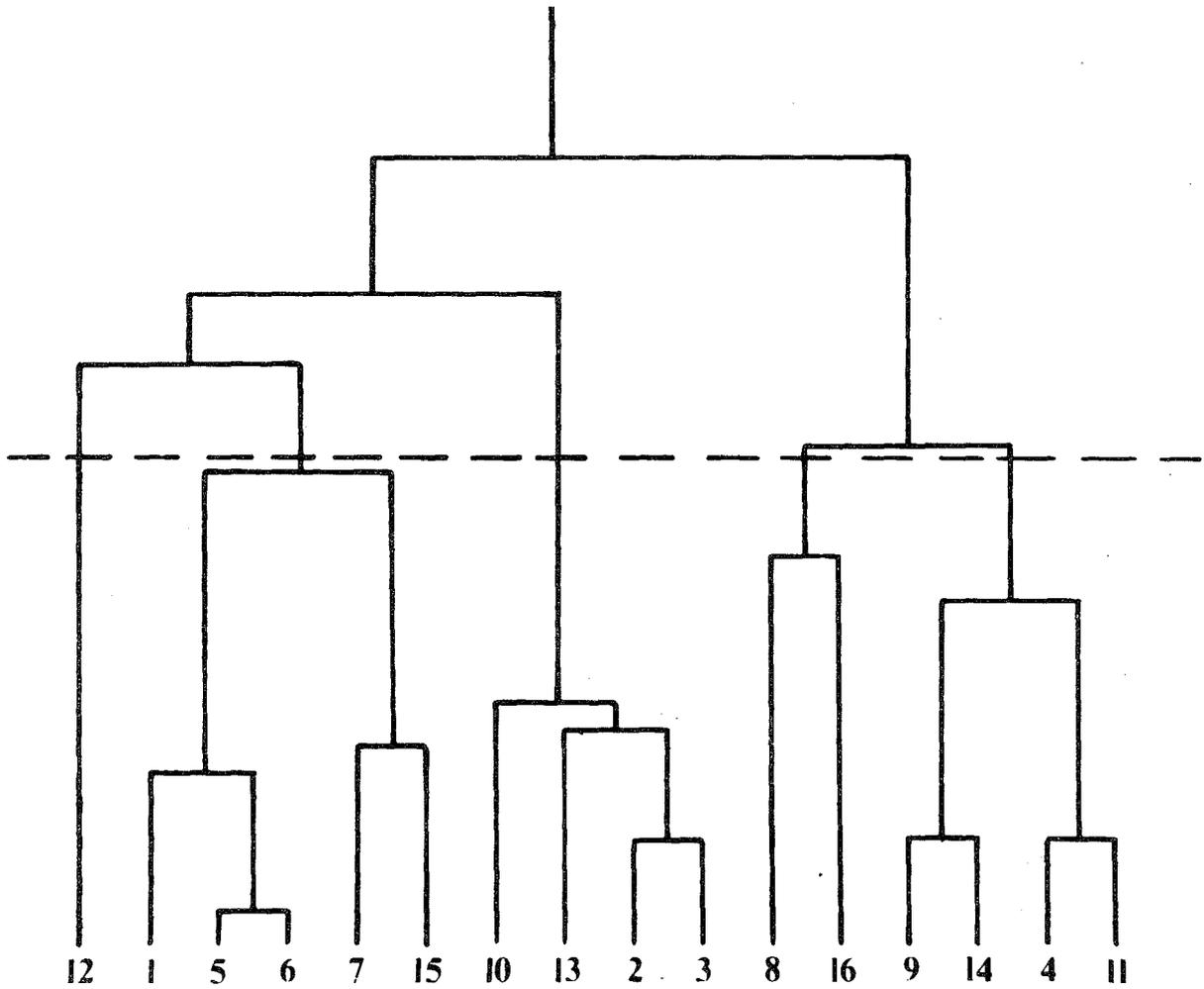


Fig. 2. Complete link dendrogram of Gotland data.

research [SYF59]; the optimal matching is shown by the starred entries. The two partitions are seen to agree in the relationships indicated between 12 of the objects; the proportion in common, 0.75, may be regarded as a measure of the agreement. The groups detected by both criteria are

(12), (1,5-7,15), (2,3,10,13) and (4,11),

Table 1. Comparison of single link and complete link partitions of the Gotland data.

		Complete link groups				
		C ₁	C ₂	C ₃	C ₄	C ₅
Single link groups	S ₁	1*	0	0	0	0
	S ₂	0	5*	0	0	0
	S ₃	0	0	4*	2	2
	S ₄	0	0	0	0	2*
	(S ₅)	0	0	0	0*	0

and one might therefore have more confidence in the existence of these groups; the other objects could be regarded as being intermediate between some of these groups, or as outliers. This procedure is also applicable to the comparison of different coverings of the same set of objects.

An alternative approach [DiS76] to the comparison of several different classifications of the same objects would be to define $\Delta(i,j)$ as the number of classifications for which objects i and j belong to different classes, and to investigate the g classes $\{C_h(k) \ (h=1, \dots, g)\}$, where $C_h(k) = \{i \mid \Delta(i,j) \leq k \ \forall j \in C_h(k)\}$. The case $k=0$ yields h "strong patterns"; for the Gotland data, these are the groups of objects corresponding to the non-empty cells of Table 1.

The illustrative example considered above compared an m -partition with an n -partition for $m = 5$ and $n = 4$. The procedure can be applied to a range of values for m and n to investigate different levels of structure in the data. The method of comparing trees, to be described in the next three subsections, attempts to carry out simultaneously such investigations at all levels of similarity.

2.2. THE INTERSECTION OF TREES

The extent to which two tree diagrams resemble one another can be measured using Sokal and Rohlf's cophenetic correlation coefficient [SoR62]. In the remainder of this section, we will be less concerned with an overall measure of agreement than in seeking to identify which parts of the classifications are contributing to the agreement. A method will be described for obtaining a maximal reduced tree which does not contradict the information contained in the two original dendrograms, *i.e.* we will seek to "prune" as few branches as possible from each of the two original trees so as to render them in some sense "equivalent". The relationships summarised in this pruned tree will retain the information on the objects about which the classifier is more confident. A disadvantage is that one would have difficulty specifying any criterion related to the objects which the reduced tree satisfies optimally; it will just be the intersection of trees, each of which will separately have optimal properties.

It is still necessary to specify what one means by "equivalent" dendrograms, and for this purpose it is useful to regard a hierarchical classification as a transformation $T : d \rightarrow \hat{d}$ of the original measure of dissimilarity $d(i,j)$ between each pair of objects i and j into a transformed dissimilarity $\hat{d}(i,j)$, where $\hat{d}(i,j)$ denotes the level of the dendrogram at which objects i and j first come into the same group. Because of the many different ways in which clustering criteria can be defined, the actual values of $\{\hat{d}(i,j)\}$ often contain little information beyond their rank ordering, and this is all that will be required in the following formulation.

Sibson [Sib72] discusses types of equivalence between two trees, T_1 and T_2 , which one may regard as being defined by the sets of transformed dissimilarities $\{\hat{d}_1(i,j)\}$ and $\{\hat{d}_2(i,j)\}$ respectively.

T_1 and T_2 are defined to be *global order equivalent* if

$$\hat{d}_1(i,j) \leq \hat{d}_1(k,l) \Leftrightarrow \hat{d}_2(i,j) \leq \hat{d}_2(k,l) \text{ for all objects } i,j,k,l.$$

T_1 and T_2 are defined to be *local order equivalent* if

$$\hat{d}_1(i,j) \leq \hat{d}_1(i,k) \Leftrightarrow \hat{d}_2(i,j) \leq \hat{d}_2(i,k) \text{ for all objects } i,j,k.$$

The aim of this section can now be stated more formally : given two dendrograms, seek to prune as few branches as possible from each of them so as to ensure that the reduced trees are global (or local) order equivalent. From experience, it seems possible that global order equivalence may be too stringent a requirement for many data sets and pairs of clustering criteria. This paper concentrates on

local order equivalence, and the next subsection describes an algorithm for seeking the optimal reduced tree.

2.3. A PRUNING ALGORITHM

Local order equivalence of two dendrograms may be stated informally as follows : consider moving up each dendrogram, starting from the same object; all objects must be encountered in the same order (including the preservation of ties in the orderings). This must hold for all starting positions. For example, consider following the branch attached to object 8 in figs. 1 and 2. In fig. 1, the objects are encountered in the order

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

where the ranks of all objects enclosed within the same pair of brackets are tied. In fig. 2, the objects are encountered in the order

8, 16, (4,9,11,14), (1-3,5-7,10,12,13,15).

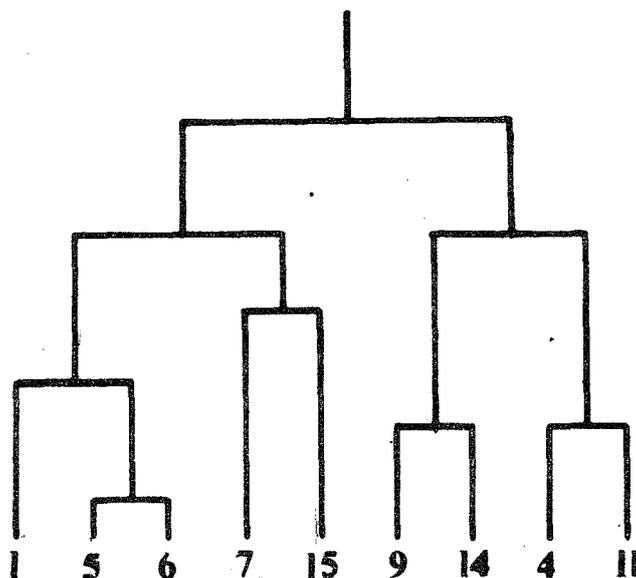


Fig. 3. A pruned tree : intersection of single link and complete link dendrograms of Gotland data.

Define α_i ($1 < i < N=16$) as the maximum agreement between the two sequences along the i^{th} branch, *i.e.* as the number of objects in the largest possible subsequence which does not contradict the orderings of either of the original sequences (including the preservation of ties). For a fuller description of this statistic and an algorithm for evaluating α_i and identifying the objects in agreement, see [Gor79]. For example, it can be verified that $\alpha_8 = 9$, with the two sequences having in common the objects

8, 16, (4,11), (1,5-7,15).

If $\sum_{i=1}^N \alpha_i = N^2$, the two dendrograms are local order equivalent. If $\sum_{i=1}^N \alpha_i < N^2$,

an iterative approximating algorithm can be used to delete one object at a time until the reduced dendrograms become equivalent. Because the set of α_i objects need not be uniquely defined, let

$$v_{ij} = \begin{cases} 1 & \text{if the } j^{\text{th}} \text{ object occurs in an optimal set of } \alpha_i \text{ objects in agreement} \\ & \text{with the } i^{\text{th}} \text{ pair of sequences,} \\ 0 & \text{otherwise.} \end{cases}$$

Define $\sigma_j = \sum_i v_{ij}$, the number of branches for which the j^{th} object occurs in an optimal subsequence. Clearly, $\sum_i \alpha_i \leq \sum_j \sigma_j$, with equality only if each sub-

sequence is uniquely defined.

The algorithm can be summarised as follows :-

1. Let M (the current size of the data set) = N .
2. For the current set of M objects, evaluate $\{\alpha_j, \sigma_j\}$.
3. If $\sum_j \alpha_j = M^2$, the reduced trees are local order equivalent. STOP.
4. If $\sum_j \alpha_j < M^2$, delete an object with minimum value of σ_j ; set $M = M-1$ and go to step 2.

The object selected for deletion at step 4 need not be uniquely defined, and while the algorithm will find a locally optimal reduced tree, there may exist larger trees satisfying the equivalence criterion. The algorithm thus shares with many other classification algorithms the property of being an approximation to an optimal solution; a complete search of all tree forms becomes rapidly less feasible as the number of objects increases.

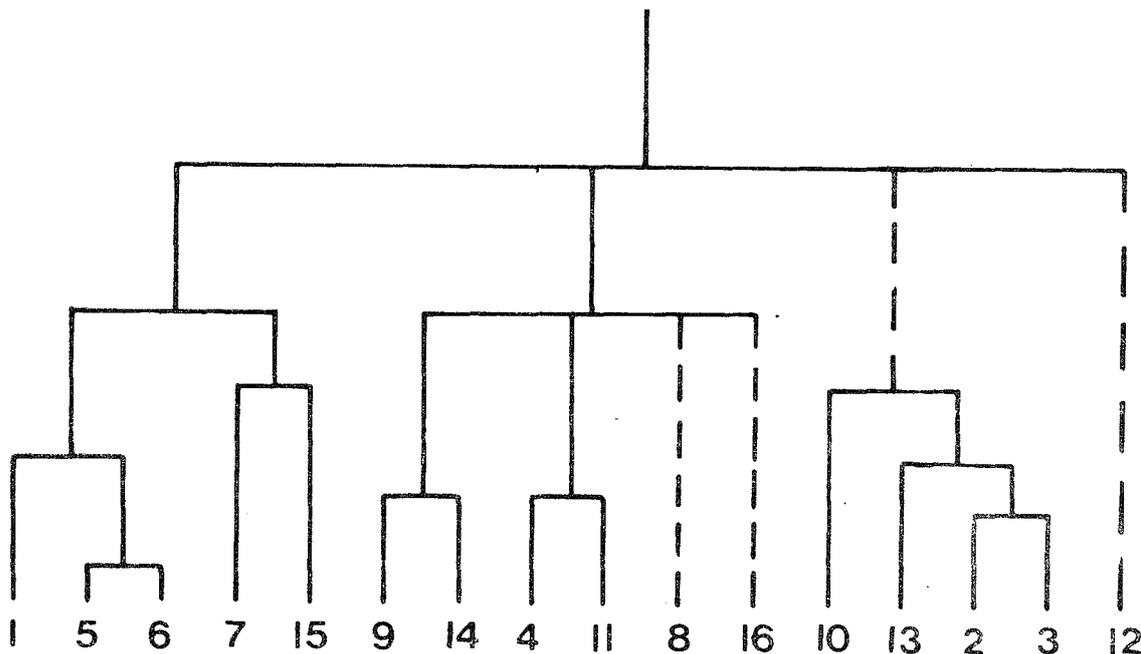


Fig. 4. A pruned and grafted tree : synthesising the information contained in the single link and complete link dendrograms of the Gotland data.

Fig. 3 gives the reduced dendrogram obtained when the above algorithm is applied to the dendrograms shown in figs. 1 and 2. There are, however, two other reduced trees containing nine objects which do not contradict the information contained in figs. 1 and 2; there is some structure contained in the branches which have been pruned, and the next subsection describes an approach which reincorporates this information into a single figure.

2.4. REGRAFTING PRUNED BRANCHES

The branches which were pruned in order to obtain the skeletal tree shown in fig. 3 are located in contradictory positions in figs. 1 and 2. In seeking to represent the relationships between all the objects in a single summary tree diagram, one approach might be to reattach each pruned branch to the skeletal tree in a position which is "average" with respect to its location in the original dendrograms; some of the algorithms described by J.A.Hartigan [Har67] could prove useful in this context. Rather than attaching a branch in a position for which neither of the original trees might offer support, a more conservative approach will be adopted

here : a branch will be attached to the reduced tree at the lowest possible level of dissimilarity for which its subsequent relationships with the skeletal set of objects are agreed between the two original trees. Low-level relationships which are common to both of the original dendrograms will be preserved in the act of grafting.

Fig. 4 shows the result of carrying out this operation : branches which were pruned then regrafted are represented by dashed lines. For example, the precise relationship between object 8 and the set of objects (4,9,11,14) is disputed in figs. 1 and 2, but its subsequent relationship with objects (1,5-7,15) is not in dispute; it joins the group (4,9,11,14) no later than the position indicated. Note also that the low-level relationships between objects 2,3,10 and 13 can be preserved in fig. 4 because they appear in both figs. 1 and 2, but the precise level at which these objects join the rest of the data set is uncertain; this feature is discussed further in the final section.

3. DISCUSSION

Throughout the paper, the work has been described in terms of comparing *two* classification schemes. The extension to comparing more than two partitions is straightforward, but it seems likely that a very efficient algorithm would be needed for the comparison of more than two dendrograms; an inefficient algorithm could get trapped in inferior local optimum solutions.

The tree-comparison method seeks very detailed relationships, and some care should be taken in the choice of clustering criteria : if two clustering criteria are best suited to uncovering markedly different structure in data, they are likely to agree closely only if there are very distinct relationships in the data set. For data for which the structure is less clear-cut, it might be advisable to use clustering criteria which are more closely related to one another [Sch70]. This would, of course, have implications for the degree of confidence with which one could claim to have detected structure in the data.

The results of using the tree-comparison method to analyse several larger data sets suggest that the method might be of most use in studying lower-level relationships between the objects. There was often close correspondence between partitions for a moderate number of groups, and a reasonable agreement about the way in which the objects within these groups were related; but it was found that the later amalgamations, between the last few groups, often took place in different ways in the two dendrograms. This observation is in accord with experience recorded in the literature [Roh70,Kru77] that clustering methods give the small-scale relationships between objects but that the large-scale relationships are better detected by ordination studies. The tree-comparison method, therefore, would seem most relevant for the examination of such small-scale relationships, when it would be desirable to identify all maximal subsets whose relationships do not contradict any of the original dendrograms. Relevant subsets could be detected by the comparison of sequences "censored" at, or before, their k^{th} member; in each sequence, all objects after the first k would receive the same rank.

Study of the results of applying the methods to a wider range of data sets would be desirable, as would a more efficient pruning algorithm, but it is believed that the methods described in this paper could make a contribution to the analysis of multivariate data.

REFERENCES

- BaH75 Baker, F.B. & Hubert, L.J. "Measuring the power of hierarchical cluster analysis", J. Amer. Statist. Assoc. 70, 1975, 31-38.
- BaH76 Baker, F.B. & Hubert, L.J. "A graph-theoretic approach to goodness-of-fit in complete link hierarchical clustering", J. Amer. Statist. Assoc. 71, 1976, 870-878.
- Bez74 Bezdek, J.C. "Numerical taxonomy with fuzzy sets", J. Math. Biology 1, 1974, 57-71.
- Bla76 Blashfield, R.K. "Questionnaire on cluster analysis software", Classification Society Bulletin 3(4), 1976, 25-42.
- Boc77 Bock, H.-H. "On tests concerning the existence of a classification", in : First International Symposium on Data Analysis and Informatics, I.R.I.A., Rocquencourt, 1977, 449-464.
- Boc79 Bock, H.-H. "Fuzzy clustering procedures", in this volume.
- Che72 Chernoff, H. "Metric considerations in cluster analysis", in : LeCam, L.M., Neyman, J. & Scott, E.L. (editors), Proc. 6th Berkeley Symp. 1, 1972, 621-629.
- Cor71 Cormack, R.M. "A review of classification", J. R. Statist. Soc. A 134, 1971, 321-367.
- Day69 Day, N.E. "Estimating the components of a mixture of normal distributions", Biometrika 56, 1969, 463-474.
- DiG77 Diday, E. & Govaert, G. "Classification automatique avec distances adaptatives", R.A.I.R.O. Informatique/Computer Science 11, 1977, 329-349.
- DiS76 Diday, E. & Simon, J.C. "Clustering analysis", in : Fu, K.S. (editor) Communication and Cybernetics 10, Springer-Verlag, Berlin etc., 1976, 47-94.
- Dom75 Domoney, D.W. "Multivariate cluster analysis : a Monte Carlo study of some basic parameters", Ph.D. thesis, University of Toronto, 1975.
- EnH69 Engelmann, L. & Hartigan, J.A. "Percentage points of a test for clusters", J. Amer. Statist. Assoc. 64, 1969, 1647-1648.
- Eve74 Everitt, B. Cluster Analysis, Heinemann Educational Books, London, 1974.
- Fra76 Frank, O. "Comparing classifications by use of the symmetric class difference", Research report 1976:3, Department of Statistics, University of Lund, 1976.
- FVN71 Fisher, L. & Van Ness, J.W. "Admissible clustering procedures", Biometrika 58, 1971, 91-104.
- GoB75 Gower, J.C. & Banfield, C.F. "Goodness-of-fit criteria in cluster analysis and their empirical distributions", in : Corsten, L.C.A. & Postelnicu, T. (editors), Proceedings of the 8th International Biometric Conference, 1975, 347-361.
- GoK54 Goodman, L.A. & Kruskal, W.H. "Measures of association for cross classifications", J. Amer. Statist. Assoc. 49, 1954, 732-764.
- Gor79 Gordon, A.D. "A measure of the agreement between rankings", Biometrika 66, 1979, 7-15.
- GrA69 Grigal, D.F. & Arneman, H.F. "Numerical classification of some forested Minnesota soils", Soil Sci. Soc. Amer. Proc. 33, 1969, 433-438.
- GPC66 Gimmingham, C.H., Pritchard, N.M. & Cormack, R.M. "Interpretation of a vegetational mosaic on limestone in the Island of Gotland", J. Ecology 54, 1966, 481-502.
- GWV68 El Gazzar, A., Watson, L., Williams, W.T. & Lance, G.N. "The taxonomy of *Salvia* : a test of two radically different numerical methods", J. Linn. Soc. (Bot.) 60, 1968, 237-250.
- Har67 Hartigan, J.A. "Representation of similarity matrices by trees", J. Amer. Statist. Assoc. 62, 1967, 1140-1158.
- Har75 Hartigan, J.A. Clustering Algorithms, Wiley, New York etc., 1975.
- Har77 Hartigan, J.A. "Distribution problems in clustering", in : Van Ryzin, J. (editor), Classification and Clustering, Academic Press, New York etc., 1977, 45-71.

- Jar70 Jardine, N. "Algorithms, methods and models in the simplification of complex data", *Computer J.* 13, 1970, 116-117.
- JaS71 Jardine, N. & Sibson, R. *Mathematical Taxonomy*, Wiley, London etc., 1971.
- Kru77 Kruskal, J. "The relationship between multidimensional scaling and clustering", in : Van Ryzin, J. (editor), *Classification and Clustering*, Academic Press, New York etc., 1977, 17-44.
- KuF75 Kuiper, F.K. & Fisher, L. "A Monte Carlo comparison of six clustering procedures", *Biometrics* 31, 1975, 777-783.
- Lin73 Ling, R.F. "A probability theory for cluster analysis", *J. Amer. Statist. Assoc.* 68, 1973, 159-164.
- Min65 Minkoff, E.C. "The effects on classification of slight alterations in numerical technique", *Syst. Zool.* 14, 1965, 196-213.
- Ran71 Rand, W.M. "Objective criteria for the evaluation of clustering methods", *J. Amer. Statist. Assoc.* 66, 1971, 846-850.
- Roh70 Rohlf, F.J. "Adaptive hierarchical clustering schemes", *Syst. Zool.* 19, 1970, 58-82.
- Sch70 Schnell, G.D. "A phenetic study of the suborder Lari (Aves) II. Phenograms, discussion and conclusions", *Syst. Zool.* 19, 1970, 264-302.
- Sib72 Sibson, R. "Order invariant methods for data analysis", *J. R. Statist. Soc. B* 34, 1972, 311-349.
- Sne77a Sneath, P.H.A. "A significance test for clusters in UPGMA phenograms obtained from squared Euclidean distances", *Classification Society Bulletin* 4(1), 1977, 2-14.
- Sne77b Sneath, P.H.A. "Cluster significance tests and their relation to measures of overlap", in : *First International Symposium on Data Analysis and Informatics*, I.R.I.A., Rocquencourt, 1977, 15-36.
- SnS73 Sneath, P.H.A. & Sokal, R.R. *Numerical Taxonomy*, Freeman, San Francisco, 1973.
- Sok77 Sokal, R.R. "Clustering and classification : background and current directions", in : Van Ryzin, J. (editor), *Classification and Clustering*, Academic Press, New York etc., 1977, 1-15.
- SoM58 Sokal, R.R. & Michener, C.D. "A statistical method for evaluating systematic relationships", *Univ. Kansas Sci. Bull.* 38, 1958, 1409-1438.
- SoR62 Sokal, R.R. & Rohlf, F.J. "The comparison of dendrograms by objective methods", *Taxon* 11, 1962, 33-40.
- SYF59 Sasieni, M., Yaspan, A. & Friedman, L. *Operations Research - Methods and Problems*, Wiley, New York, 1959.
- Tuk77 Tukey, J.W. *Exploratory Data Analysis*, Addison-Wesley, Reading Massachusetts etc., 1977.
- VNe73 Van Ness, J.W. "Admissible clustering procedures", *Biometrika* 60, 1973, 422-424.
- War63 Ward, J.H. "Hierarchical grouping to optimize an objective function", *J. Amer. Statist. Assoc.* 58, 1963, 236-244.
- WiD65 Williams, W.T. & Dale, M.B. "Fundamental problems in numerical taxonomy", in: Preston, R.D. (editor), *Advances in Botanical Research*, volume 2, Academic Press, London, 1965, 35-68.
- Wis69 Wishart, D. "Mode analysis : a generalisation of nearest neighbour which reduces chaining effects", in : Cole, A.J. (editor), *Numerical Taxonomy*, Academic Press, London etc., 1969, 282-311.
- Wol70 Wolfe, J.H. "Pattern clustering by multivariate mixture analysis", *Multivar. Behav. Res.* 5, 1970, 329-350.
- WWL66 Watson, L., Williams, W.T. & Lance, G.N. "Angiosperm taxonomy : a comparative study of some novel numerical techniques", *J. Linn. Soc. (Bot.)* 59, 1966, 491-501.