Journal of Computational and Applied Mathematics 20 (1987) 53–65
North-Holland

Silhouettes: a graphical aid to the interpretation and validation of cluster analysis

Peter J. ROUSSEEUW
University of Fribourg, ISES, CH-1700 Fribourg, Switzerland

Received 13 June 1986
Revised 27 November 1986

Abstract: A new graphical display is proposed for partitioning techniques. Each cluster is represented by a so-called *silhouette*, which is based on the comparison of its tightness and separation. This silhouette shows which objects lie well within their cluster, and which ones are merely somewhere in between clusters. The entire clustering is displayed by combining the silhouettes into a single plot, allowing an appreciation of the relative quality of the clusters and an overview of the data configuration. The average silhouette width provides an evaluation of clustering validity, and might be used to select an 'appropriate' number of clusters.

Keywords: Graphical display, cluster analysis, clustering validity, classification.

1. The need for graphical displays

There are many algorithms for partitioning a set of objects into k clusters, such as the k-means method [6,9,13] and the k-median approach [20]. The result of such a partitioning technique is a list of clusters with their objects, which is not as visually appealing as the dendrograms of hierarchical methods. It is hoped that the graphical display introduced in Section 2 will contribute to the interpretation of cluster analysis results, as illustrated by the examples of Section 3. In Section 4, some other displays will be described.

Suppose there are n objects to be clustered, which may be persons, flowers, cases, statistical variables, or whatever. Clustering algorithms mainly operate on two frequently used input data structures (see [18, Chapters 1 and 2]). The first method is to represent the objects by means of a collection of measurements or attributes, such as height, weight, sex, color, and so on. In Tucker's [19] terminology such an objects by attributes matrix is called two-mode, since the row and column entities are different. When the measurements are on an interval scale, one can compute the Euclidean distance $d(i, j)$ between any objects i and j.

This leads us to the second data structure, namely a collection of proximities which must be available for all pairs of objects. This corresponds to a one-mode matrix, since the row and column entities are the same set of objects. We shall consider two types of proximities: dissimilarities (which measure how far away two objects are from each other) and similarities

Table 1
Dissimilarities between twelve countries, obtained by averaging the result of a survey among political science students

Country	Dissimilarities to other countries										
	BEL	BRA	CHI	CUB	EGY	FRA	IND	ISR	USA	USS	YUG
BRA	5.58										
CHI	7.00	6.50									
CUB	7.08	7.00	3.83								
EGY	4.83	5.08	8.17	5.83							
FRA	2.17	5.75	6.67	6.92	4.92						
IND	6.42	5.00	5.58	6.00	4.67	6.42					
ISR	3.42	5.50	6.42	6.42	5.00	3.92	6.17				
USA	2.50	4.92	6.25	7.33	4.50	2.25	6.33	2.75			
USS	6.08	6.67	4.25	2.67	6.00	6.17	6.17	6.92	6.17		
YUG	5.25	6.83	4.50	3.75	5.75	5.42	6.08	5.83	6.67	3.67	
ZAI	4.75	3.00	6.08	6.67	5.00	5.58	4.83	6.17	5.67	6.50	6.92

(which measure how much they resemble each other). In this paper we shall assume proximities on a *ratio scale*, such as the Euclidean distances mentioned. Also, most of the discussion will be concentrated on dissimilarities, but the formulas for similarities (with analogous interpretation) are also presented.

Let us consider a real data set consisting of dissimilarities. A questionnaire was distributed in a political science class, asking the students to provide ratings of perceived positive dissimilarity between all distinct pairs of twelve countries on a scale from 1 to 9. This is a replication of a well-known experiment of Wish [23] with the students' own country (Belgium) included. The countries were (in alphabetical order): Belgium (BEL), Brasil (BRA), mainland China (CHI), Cuba (CUB), Egypt (EGY), France (FRA), India (IND), Israel (ISR), United States (USA), USSR (USS), Yugoslavia (YUG), and Zaire (ZAI). It was imposed that the dissimilarities had to be symmetric, and the dissimilarity of a country to itself was not recorded. The final dissimilarity coefficients listed in Table 1 were obtained by taking the averages of the values given by the students. By construction, Table 1 is only a triangular half matrix.

Still, Table 1 contains 66 numbers, making it hard to perceive the structure of the data by the naked eye. To obtain a better insight, one can partition the countries into k clusters. For instance, one can use the k-median method as described by Massart, Plastria and Kaufman [14], in which k representative objects are selected so as to minimize

$$\frac{1}{n}\sum_{i=1}^{n} d(i, m(i))$$

where $d(i, m(i))$ is the dissimilarity of object i to the nearest representative object, denoted by $m(i)$. Like the k-means method, this algorithm tries to find roughly spherical clusters. This yields the following result for $k = 2$:

cluster 1: BEL, BRA, EGY, FRA, ISR, USA, ZAI
cluster 2: CHI, CUB, IND, USS, YUG

and gives the following clustering for $k = 3$:

cluster 1: BEL, EGY, FRA, ISR, USA
cluster 2: BRA, IND, ZAI
cluster 3: CHI, CUB, USS, YUG

However, this does not tell whether these partitions reflect a clustering structure actually present in the data, or if we have merely partitioned our objects into some artificial groups. Indeed, such clustering methods always come up with k groups, whatever the data are like. At this stage, we are left with many questions. Are the clusters of a high quality (that is, are the 'within' dissimilarities small when compared to the 'between' dissimilarities)? Which objects appear to be well-classified, which ones are misclassified, and which ones lie in between clusters? What is the overall structure of the data like? Can we obtain an idea about the number of 'natural' clusters that are really present? These questions are difficult, and we feel that the existing displays answer them only partially. It is hoped that the silhouettes introduced in the next section will provide the user with additional guidance.

2. Construction of silhouettes

The silhouettes constructed below are useful when the proximities are on a ratio scale (as in the case of Euclidean distances) and when one is seeking compact and clearly separated clusters. Indeed, the definition makes use of average proximities as in the case of group average linkage, which is known to work best in a situation with roughly spherical clusters.

In order to construct silhouettes, we only need two things: the partition we have obtained (by the application of some clustering technique) and the collection of all proximities between objects. For each object i we will introduce a certain value $s(i)$, and then these numbers are combined into a plot.

Let us first define the numbers $s(i)$ in the case of dissimilarities. Take any object i in the data set, and denote by A the cluster to which it has been assigned. (For a concrete illustration, see Fig. 1). When cluster A contains other objects apart from i, then we can compute

$$a(i) = \text{average dissimilarity of } i \text{ to all other objects of } A.$$

In Fig. 1, this is the average length of all lines within A. Let us now consider any cluster C which is different from A, and compute

$$d(i, C) = \text{average dissimilarity of } i \text{ to all objects of } C.$$

In Fig. 1, this is the average length of all lines going from i to C. After computing $d(i, C)$ for all clusters $C \neq A$, we select the smallest of those numbers and denote it by

$$b(i) = \min_{C \neq A} d(i, C).$$

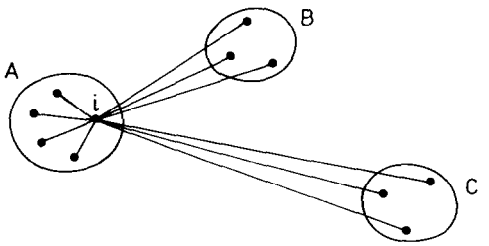

Fig. 1. An illustration of the elements involved in the computation of $s(i)$, where the object i belongs to cluster A.

The cluster B for which this minimum is attained (that is, $d(i, B) = b(i)$) we call the *neighbor* of object i. This is like the second-best choice for object i: if it could not be accommodated into cluster A, which cluster B would be the closest competitor? In Fig. 1, cluster B indeed appears to be 'closest' (on the average) to object i, when A itself is discarded. Therefore, it is very useful to know the neighbor of each object in the data set. Note that the construction of $b(i)$ depends on the availability of other clusters apart from A, so we have to assume throughout this paper that the number of clusters k is more than one.

The number $s(i)$ is obtained by combining $a(i)$ and $b(i)$ as follows:

$$s(i) = \begin{cases} 1 - a(i)/b(i) & \text{if } a(i) < b(i), \\ 0 & \text{if } a(i) = b(i), \\ b(i)/a(i) - 1 & \text{if } a(i) > b(i). \end{cases}$$

It is even possible to write this in one formula:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}.$$

When cluster A contains only a single object it is unclear how $a(i)$ should be defined, and then we simply set $s(i)$ equal to zero. This choice is of course arbitrary, but a value of zero appears to be most neutral. Indeed, from the above definition we easily see that

$$-1 \leqslant s(i) \leqslant 1$$

for each object i.

Note that the $s(i)$ defined above remains invariant when all the original dissimilarities are multiplied by a positive constant, but that an additive constant is not allowed. This explains why we have explicitly assumed that the dissimilarities were on a *ratio scale*, which means that a dissimilarity of 6 may be considered twice as large as a dissimilarity of 3. For instance, Euclidean distances are on a ratio scale.

To strengthen our intuition about the meaning of $s(i)$, let us look at a few extreme situations. When $s(i)$ is at its largest (that is, $s(i)$ close to 1) this implies that the 'within' dissimilarity $a(i)$ is much smaller than the smallest 'between' dissimilarity $b(i)$. Therefore, we can say that i is 'well-clustered', as there appears to be little doubt that i has been assigned to a very appropriate cluster: the second-best choice (B) is not nearly as close as the actual choice (A).

A different situation occurs when $s(i)$ is about zero. Then $a(i)$ and $b(i)$ are approximately equal, and hence it is not clear at all whether i should have been assigned to either A or B. Object i lies equally far away from both, so it can be considered as an 'intermediate case'.

The worst situation takes place when $s(i)$ is close to -1. Then $a(i)$ is much larger than $b(i)$, so i lies on the average much closer to B than to A. Therefore it would have seemed much more natural to assign object i to cluster B, so we can almost conclude that this object has been 'misclassified'.

To conclude, $s(i)$ measures how well object i matches the clustering at hand (that is, how well it has been classified). In the special case where there are only two clusters ($k = 2$), we note that shifting object i from one cluster to the other will convert $s(i)$ to $-s(i)$.

In case the data consist of *similarities*, which also must be on a ratio scale, some small modifications must be made. We now define $a'(i)$ and $d'(i, C)$ as the corresponding average similarities, and put

$$b'(i) = \underset{C \neq A}{\text{maximum}}\, d'(i, C).$$

The numbers s(i) given by

$$s(i) = \begin{cases} 1 - b'(i)/a'(i) & \text{if } a'(i) > b'(i), \\ 0 & \text{if } a'(i) = b'(i), \\ a'(i)/b'(i) - 1 & \text{if } a'(i) < b'(i), \end{cases}$$

may then be interpreted in the same way as before.

Having computed the quantities $s(i)$ from either similarities or dissimilarities, we can now construct the graphical display. The silhouette of A is a plot of the $s(i)$, ranked in decreasing order, for all objects i in A. On a line printer, we represent $s(i)$ by a row of asterisks, the length of which is proportional to $s(i)$. Therefore, the silhouette shows which objects lie well within their cluster, and which ones are merely somewhere in between clusters. A wide silhouette indicates large $s(i)$ values, and hence a pronounced cluster. The other dimension of a silhouette is its height, which simply equals the number of objects in A.

In order to obtain an overview, the silhouettes of the different clusters are printed below each other. In this way the entire clustering can be displayed by means of a single plot, which enables us to distinguish 'clear-cut' clusters from 'weak' ones.

Figure 2 shows the silhouettes for the clustering into $k = 2$ clusters of the twelve countries data mentioned above. The first silhouette is higher than the second, because the first cluster contains seven objects and the second only five. The leftmost column (CLU) contains the index of each cluster (1 and 2), and the second column (NEIG) gives the neighbor of each object. The third column lists the numbers $s(i)$, and the fourth identifies each object i by means of its three-character label. Above and below the plot we find scales going from 0.00 to 1.00 with steps of size 0.04 (to be read vertically).

Our first impression is that both silhouettes in Fig. 2 are rather narrow, which indicates a relatively weak clustering structure. The first cluster consists of Western industrialized countries and developing countries. In this cluster, USA possesses the largest $s(i)$ which means that it was classified with the least amount of doubt. The three developing countries (BRA, EGY, ZAI) are listed at the end because they have smaller $s(i)$ values than the four capitalist nations. The second cluster consists of four Communist countries and India. The $s(i)$ values of the Communist countries are comparable to those of the Western countries in cluster 1. However, for India we find $s(i) = -0.04$, and thus this country is an intermediate case lying far from both clusters. Although the k-median algorithm assigns India to cluster 2, one could also make a case that India should belong to cluster 1 because $s(i)$ is even slightly negative; moving it to the other cluster would yield $s(i) = +0.04$.

Figure 3 contains the silhouette plot of the same data, but now partitioned into 3 clusters (also by means of the k-median method). Cluster 1 consists of the four Western countries plus Egypt, but we see that the $s(i)$ value of the latter nation is approximately zero, which means that it holds an intermediate position between clusters 1 and 2 (because we see in the second column that the neighbor of Egypt is cluster 2). The second cluster contains the remaining developing

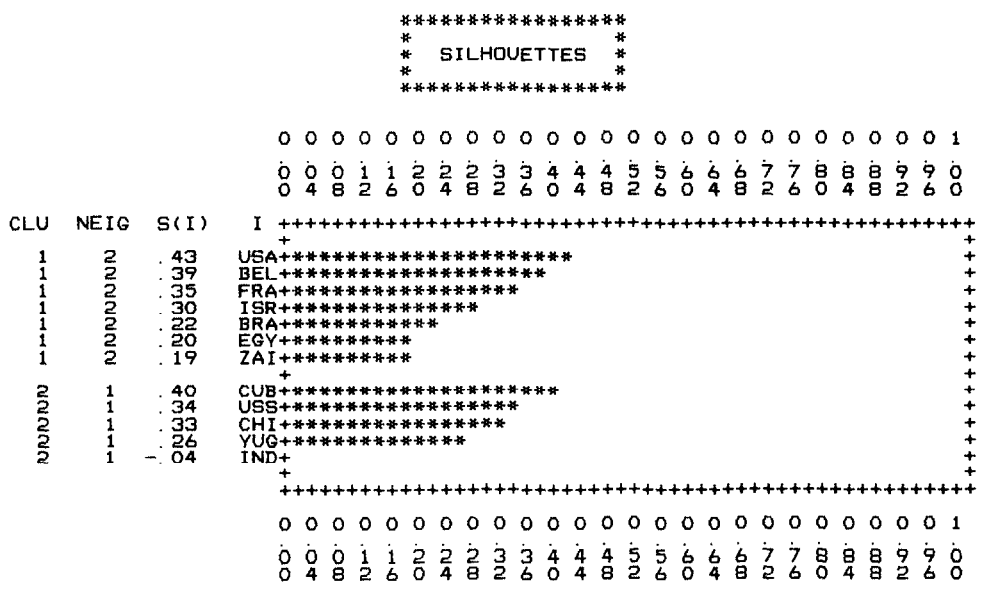

Fig. 2. Silhouettes of a clustering with $k = 2$ of the twelve countries data of Table 1.

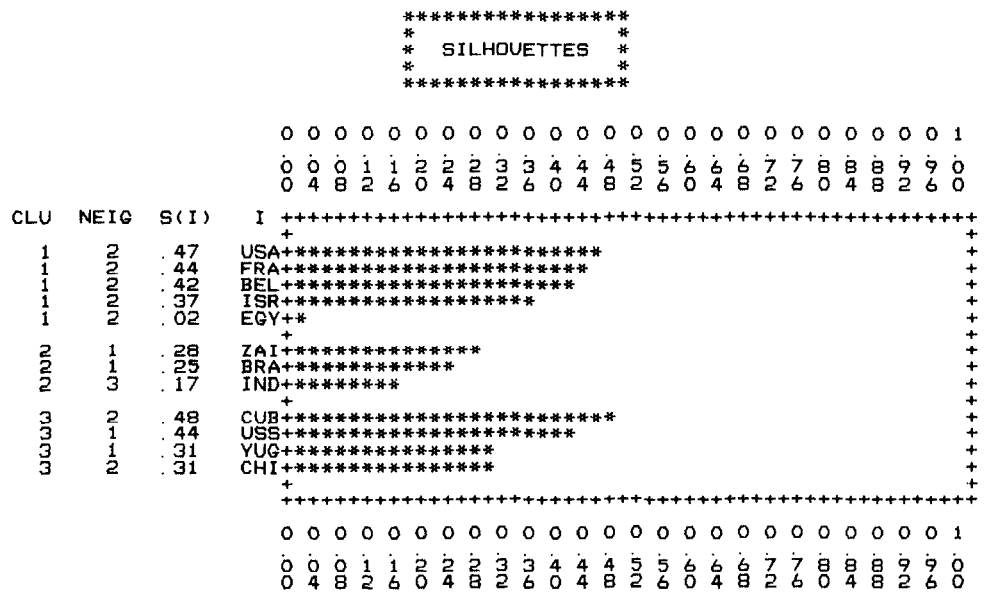

Fig. 3. Silhouettes of a clustering with $k = 3$ of the twelve countries data.

countries, and the third one consists of the four Communist nations in the survey. The silhouettes of the capitalist and the Communist countries are now wider than in Fig. 2, which means that these clusters are slightly more pronounced. On the other hand, the second cluster does not score so highly. In the first cluster, all objects have cluster 2 for their neighbor. In the second cluster, it appears that Zaire and Brasil are more inclined towards the Western countries because their neighbor is cluster 1, whereas India (which has a smaller value of $s(i)$) seems to be closer to cluster 3. In the third cluster there is also a dichotomy: USSR and Yugoslavia seem to have some resemblance to the Western countries, whereas Cuba and China appear to be closer to the developing ones.

Silhouettes offer the advantage that they only depend on the actual partition of the objects, and not on the clustering algorithm that was used to obtain it. As a consequence, silhouettes could be used to improve the results of cluster analysis (for instance by moving an object with negative $s(i)$ to its neighbor), or to compare the output of different clustering algorithms applied to the same data.

However, we think the main usefulness of silhouettes lies in the interpretation and validation of cluster analysis results. Let us consider a heuristic argument relating silhouettes to the number of clusters. Suppose the data set consists of some dense clusters which are far away from each other, but that we have set k too low. In this case, most clustering algorithms will combine some natural clusters in order to reduce the total number of groups to the specified value of k. Fortunately, the silhouette plot will often expose such artificial fusions. Indeed, joining different clusters will lead to large 'within' dissimilarities and hence to large $a(i)$, resulting in small $s(i)$ values for the objects in such a conglomerate, yielding a narrow silhouette. ('Narrow' is meant in a relative sense, because comparisons are made across several values of k.)

On the other hand, suppose that we have set k too high. Then some natural clusters have to be divided in an artificial way, in order to conform to the specified number of groups. However, these artificial fragments will typically also show up through their narrow silhouettes. Indeed, the objects in such a fragment are on the average very close to the remaining part(s) of their natural cluster, and hence the 'between' dissimilarities $b(i)$ will become very small, which also results in small $s(i)$ values.

This heuristic reasoning implies that the silhouettes should look best for a 'natural' value of k. Therefore, we want the silhouettes to be as wide (or as dark) as possible. For each cluster, we can define the *average silhouette width* as the average of the $s(i)$ for all objects i belonging to that cluster. This allows us to distinguish 'clear-cut' from 'weak' clusters in the same plot: clusters with a larger average silhouette width are more pronounced. In Fig. 3, we see that the average silhouette width of the second cluster is only 0.24, whereas the first and third cluster attain higher values.

We can also consider the *overall average silhouette width* for the entire plot, which is simply the average of the $s(i)$ for all objects i in the whole data set. In Fig. 2 this yields 0.28, and in Fig. 3 we obtain 0.33. In general, each value of k will yield a different overall average silhouette width $\bar{s}(k)$. One way to choose k 'appropriately' is to select that value of k for which $\bar{s}(k)$ is a large as possible. For the twelve countries data, the compution of $\bar{s}(k)$ for the k-median partitions corresponding to $k = 2, \ldots, 12$ yields $k = 3$ as the best choice.

3. Examples

Let us now look at some rather extreme examples to obtain a better feeling for the meaning of silhouettes. First suppose we have eight points which are divided over some very tight clusters, far away from each other. For instance, assume that five objects coincide with one geometrical point, and the remaining three coincide with another geometrical point at a large distance from the first. This is an extremely sharp clustering structure, which should be recovered by any reasonable algorithm when $k = 2$. The resulting silhouette plot in Fig. 4 is as dark as possible. All $s(i)$ equal 1.00, so the overall average silhouette width attains its maximal value 1.00 (therefore, $k = 2$ is the best choice). In general, a very large overall average silhouette width can be taken to mean that the clustering algorithm has discovered a very strong clustering structure.

However, one must be a little careful with the interpretation of the last statement. For instance, a situation where the data set contains one far outlier is also an example of a strong clustering structure. Indeed, when the outlier is far enough, the other data look like a tight cluster by comparison. We may expect our clustering algorithm to provide the following picture for $k = 2$: one cluster containing just the outlier, and the other cluster consisting of the bulk of the data. For instance, suppose seven objects have zero dissimilarities to each other, and all have large dissimilarities to the eighth object, which is an outlier. The resulting silhouette plot looks like Fig. 5, clearly separating object 8 from the rest. Because cluster 2 contains only a single object, its $s(i)$ was put equal to zero by convention. In order to mark such singleton clusters in a more distinctive way, John Tukey (personal communication) suggested printing the number 1 in the rightmost column of the plot. The overall average silhouette width equals 0.88, which is very high (in general, one obtains $(n-1)/n$ where n is the total number of objects). Therefore, one should never merely accept a high overall average silhouette width at its face value, but also look

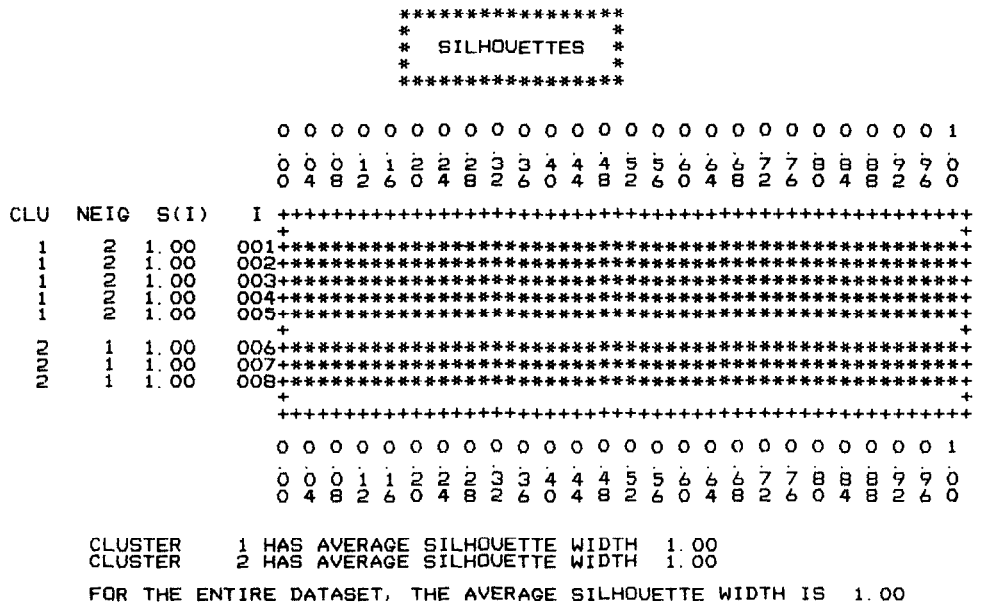

Fig. 4. Silhouettes of an example where eight points are divided over two very tight clusters, for $k = 2$.

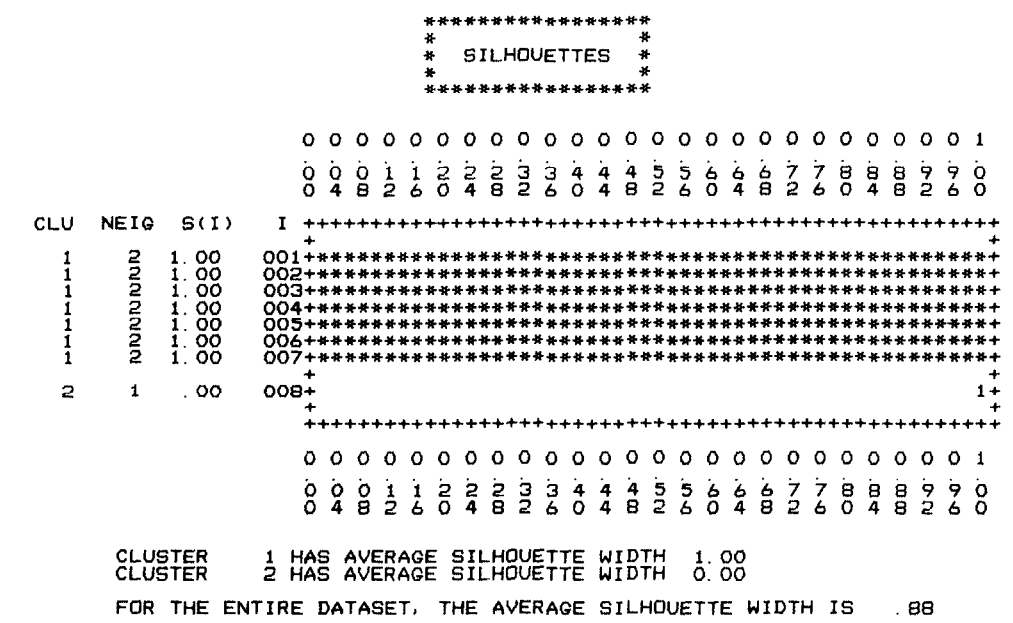

Fig. 5. Silhouettes of a data set containing a distant outlier, for $k = 2$.

at the graphical output itself to find out what caused it. Depending on the subject matter and the task at hand, one might want to put the outlier aside for further investigation and run the clustering algorithm again on the remaining data.

When the clustering algorithm does not succeed in finding any 'natural' clustering, the overall average silhouette width tends to become very low. An extreme case is when all dissimilarities between pairs of objects equal the same positive constant, so all non-diagonal entries of the dissimilarity matrix are identical. (In the case of Euclidean distances, this happens with the n vertices of a regular simplex in $n-1$ dimensions.) In such a situation no clustering is more natural than any other, so there is a total absence of clustering structure. Whatever the value of k, and whatever clustering algorithm is used, all $s(i)$ will be zero (as well as the overall average silhouette width $\bar{s}(k)$) and the silhouette plot stays completely empty. In actual applications one sometimes does encounter cases where even the largest value of $\bar{s}(k)$ is rather small, and then one might better seek an alternative to clustering as a model for and method of data analysis.

Figure 6 shows a plot of the well-known Ruspini data [15]. This data set is composed of 75 points, and was originally used by Ruspini to illustrate fuzzy clustering techniques. (The actual coordinates were communicated to me by G. Libert). The points make up four groups *A, B, C, D* as indicated in the plot. Because this example is two-dimensional, we can now compare the silhouettes to the structure that we perceive by the naked eye. Figure 7 contains the silhouette plots of the k-median clusterings with $k = 2, \ldots, 6$ (making use of Euclidean distance).

In the case of $k = 2$, one cluster is formed as the union of A with D, whereas the second combines B and C. As we saw before, artificial fusions are penalized by narrow silhouettes. For $k = 3$ we see that clusters B and C are found, but A and D still stick together. The corresponding silhouette plot shows clearly that both B and C are more pronounced than the union of A with D. For $k = 4$ the 'right' solution is found, which leads to four silhouettes of

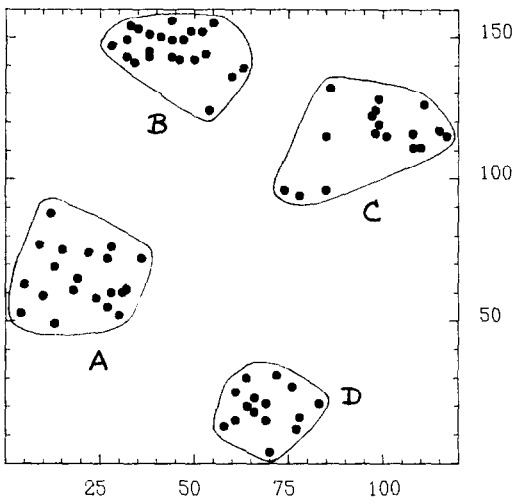

Fig. 6. Plot of the two-dimensional data set of Ruspini. The loops are merely drawn to indicate four groups of points.

about the same good quality. When $k = 5$ is imposed, the algorithm splits C into two parts. The second part contains the three 'lowest' points of C (as viewed in Fig. 6), that is, the three points of C with smallest y-coordinates. This trio has a rather prominent silhouette, and indeed some people consider it as a genuine cluster (see [4]). However, the silhouette of the major part of C becomes somewhat less wide, because this cluster is not so well separated from the three-point one (indeed, one object has an $s(i)$ value of about zero because it lies rather close to the three-point cluster and therefore holds an intermediate position). The last case ($k = 6$) leads to a more dramatic effect: cluster A is being split up in an artificial way, and consequently both parts obtain narrow silhouettes. For $k = 7, \ldots, 75$ the results are still worse.

In conclusion, the silhouette plots tell us that a division into $k = 4$ clusters is probably most natural. Indeed, the overall average silhouette width $\bar{s}(k)$ is largest for $k = 4$ (even if one tries all values of k ranging from 2 to 75). The second best $\bar{s}(k)$ is attained for $k = 5$, and the silhouette plot shows us the advantages and disadvantages of the corresponding clustering.

4. Related graphical displays

Let us now look at some existing displays for cluster analysis. In the literature most of the emphasis has been on hierarchical analysis, the results of which can be described by means of trees or dendrograms. Also other graphical tools have been developed for summarizing hierarchical results, such as the diagrams of Ward [22], Johnson [10], Kruskal and Landwehr [12], Kent [11], and others. Like the silhouette plot, the latter displays only require a line printer. On the other hand, one can also construct a two-dimensional representation of the data (e.g. by means of multidimensional scaling) and then indicate nested clusters by drawing loops around the objects [16].

In the case of partitioning, relatively few displays have been considered. There are however some examples, like shading of the dissimilarity matrix (which has also been used in connection

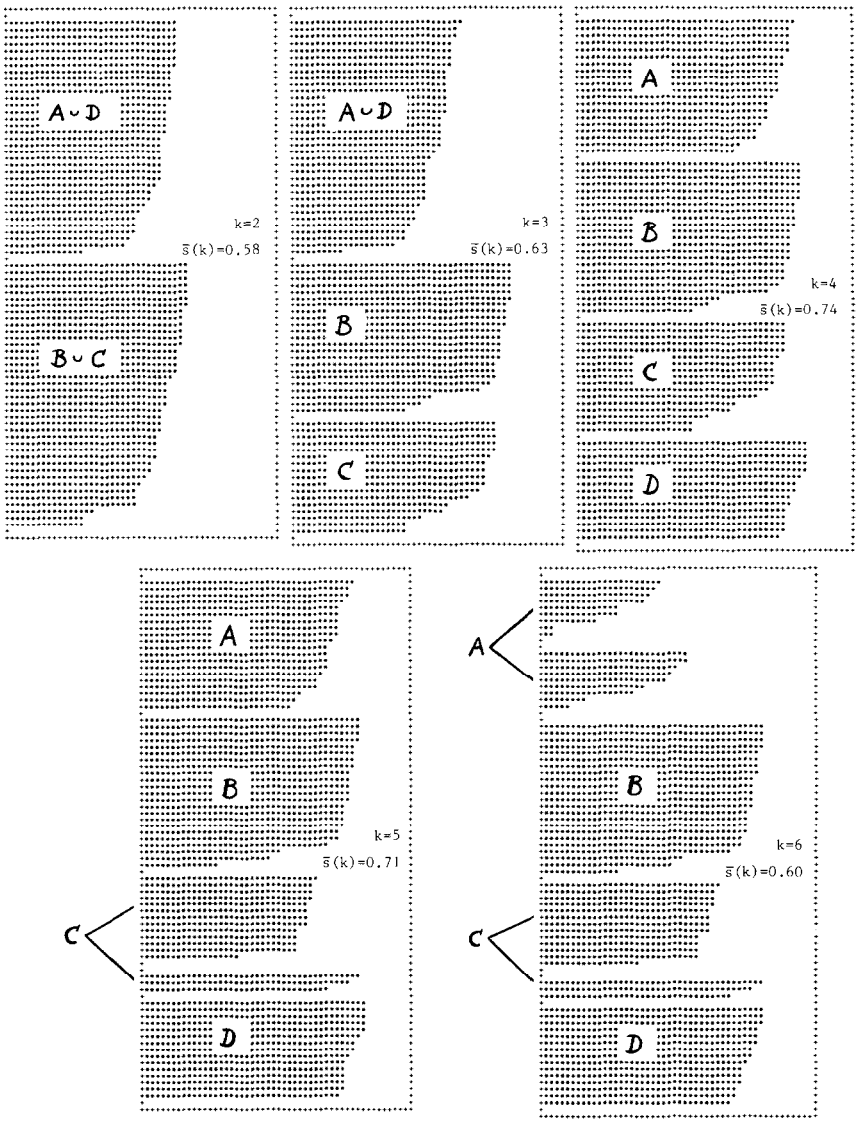

Fig. 7. Silhouette plots of the Ruspini data, for k ranging from 2 to 6.

with hierarchical methods). In the lower triangular matrix of dissimilarities the numbers are replaced by symbols, the darkness of which depends on the magnitude of the original entries. When the objects are in some random order this does not reveal much, but when the objects are ranked according to a clustering then a structure becomes visible (see [17, p. 110] for an interesting example from biology). A display like this can be generated automatically and drawn with a line printer. Recently, Gale, Halperin, and Costanzo [7] proposed a more refined version based on unclassed chloropeth mapping.

Wainer [21] gives a survey of methods for displaying multivariate data, in which each object is represented by an icon (such as a polygon or a face) made up of parts which vary in size or shape

with the measured attributes. Also some techniques are mentioned for allowing tables to better communicate the data structure, such as rounding, reordering, and blocking.

Another approach is the taxometric map [2] which portrays a cluster as a circle, the diameter of which is proportional to the diameter of the cluster. One tries to place the circles in the map in such a way that the distances between them are proportional to the dissimilarities between the corresponding clusters (these inter-cluster dissimilarities may be defined in different ways). For the pairs of clusters for which this is possible, a straight full line is drawn between the circles. When the distance on the map would be too large the line is partly dashed, and when it is too small a V-shaped line of correct length is drawn. A more extensive discussion of taxometric maps is provided by Everitt [5], who gives an example. Taxometric maps are complementary to silhouettes, because they depict the relationship between clusters but not the position of the individual objects within those clusters.

Many variants of silhouette plots can be thought of, for instance by using more sophisticated plotting devices instead of a line printer. One could also plot the negative $s(i)$ (we have not done so in the present examples because only a few slightly negative $s(i)$ occurred, and we wanted to use the width of the plot to its fullest advantage). J. Tukey (personal communication) also suggested ways to modify the ordering of objects and clusters in the plot.

The overall silhouette width $\bar{s}$ is not the only combination of the individual $s(i)$ that could be used for choosing a 'most appropriate' value of k: one might use squares; medians, and so on. Like most validity coefficients, also $\bar{s}$ could be used as an objective function for the clustering itself (that is, one might want to find a clustering which maximizes $\bar{s}$). Another idea would be to define $s(i)$ simply by $1 - a(i)/b(i)$, so $s(i)$ could become much smaller than -1, hereby penalizing misclassified points to a larger extent. Also, when the clustering method is based on the construction of centroids or representative objects (one for each cluster), one could use the dissimilarities to these centroids, which takes fewer calculations than computing $a(i)$ and $d(i, C)$. Such a display would, however, depend on the clustering algorithm.

Another display is the distance graph used by Cohen, Gnanadesikan, Kettenring, and Landwehr [3] and Gnanadesikan, Kettenring, and Landwehr [8]. For each cluster centroid, they plot the distances of every entity from that centroid (the symbol plotted is the cluster to which the entity was assigned). Although this display gives an indication of the internal cohesiveness of a cluster, it does not allow the study of individual objects because they remain anonymous in each list. For a recent sociological application, see Andes [1].

Acknowledgement

The author is grateful to L. Kaufman and J.W. Tukey for helpful comments improving the presentation of this article.

References

[1] N. Andes, Application of validity techniques on a hierarchical cluster solution using U.S. occupations, paper presented at the Fourth European Meeting of the Psychometric Society and the Classification Societies, Cambridge, United Kingdom, 2–5 July, 1985.

[2] J.W. Carmichael and P.N.A. Sneath, Taxometric maps, *Systematic Zoology* **18** (1969) 402–415.
[3] A. Cohen, R. Gnanadesikan, J. Kettenring and J.M. Landwehr, Methodological developments in some applications of clustering, in: P.R. Krishnaiah, Ed., *Applications of Statistics* (North-Holland, Amsterdam, 1977) 141–162.
[4] M. Delattre and P. Hansen, Bicriterion cluster analysis, *IEEE Trans. Pattern Anal. Mach. Intelligence* **2** (1980) 277–291.
[5] B. Everitt, *Graphical Techniques for Multivariate Data* (Heinemann, London, 1978).
[6] E.W. Forgy, Cluster analysis of multivariate data: efficiency versus interpretability of classifications (abstract), *Biometrics* 21 (1965) 768.
[7] N. Gale, W.C. Halperin and C.M. Costanzo, Unclassed matrix shading and optimal ordering in hierarchical cluster analysis, *J. Classification* **1** (1984) 75–92.
[8] R. Gnanadesikan, J. Kettenring and J.M. Landwehr, Interpreting and assessing the results of cluster analyses, *Bull. Internat. Statist. Inst.* **47** (1977) 451–463.
[9] R.C. Jancey, Multidimensional group analysis, *Australian J. Botany* **14** (1966) 127–130.
[10] S.C. Johnson, Hierarchical clustering schemes, *Psychometrika* 32 (1967) 241–254.
[11] Ph. Kent, A comment on icicles, *Amer. Statist.* 38 (1984) 162–163.
[12] J.B. Kruskal and J.M. Landwehr, Icicle plots: better displays for hierarchical clustering, *Amer. Statist.* 37 (1983) 162–168.
[13] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: L. Le Cam and J. Neyman, Eds., *Proceedings 5th Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, 1967) 281–297.
[14] D.L. Massart, F. Plastria and L. Kaufman, Non-hierarchical clustering with MASLOC, *Pattern Recognition* **16** (1983) 507–516.
[15] H.R. Ruspini, Numerical methods for fuzzy clustering, *Information Sciences* **2** (1970) 319–350.
[16] R.N. Shepard, Representation of structure in similarity data: problems and prospects, *Psychometrika* **39** (1974) 373–421.
[17] R.R. Sokal, Numerical taxonomy, *Scientific American* (December 1966) 106–116.
[18] H. Spaeth, *Cluster Analysis Algorithms* (Ellis Horwood, Chichester, 1980).
[19] L.R. Tucker, The extension of factor analysis to three-dimensional matrices, in: N. Frederiksen and H. Gulliksen, Eds., *Contributions to Mathematical Psychology* (Holt, Rinehart and Winston, New York, 1964).
[20] H.D. Vinod, Integer programming and the theory of grouping, *J. Amer. Statist. Assoc.* **64** (1969) 506–519.
[21] H. Wainer, On multivariate display, in: M.H. Rivzi, J.S. Rustagi and D. Siegmund, Eds., *Recent Advances in Statistics* (Academic Press, New York, 1983) 469–508.
[22] J.H. Ward, Hierarchical grouping to optimize an objective function, *J. Amer. Statist. Assoc.* **58** (1963) 236–244.
[23] M. Wish, Comparisons among multidimensional structures of nations based on different measures of subjective similarity, *General Systems* **15** (1970) 55–65.