3. HICLUS (HIerarchical CLUStering)

1. OVERVIEW

Concisely: HICLUS (HIerarchical CLUStering) provides analysis of two-way (dis)similarity data by means of a hierarchical clustering scheme using a monotonic transformation of the data.

Since HICLUS does not employ a spatial representation, the Carroll-Arabie (1979) classification is not useful in describing the program.

1.1 ORIGIN, VERSIONS AND ACRONYMS

HICLUS was originally programmed by Johnson (1967) following work by Ward (1963). The present program is based on the original Bell Laboratories version of the program.

1.2 HICLUS IN BRIEF

The method of hierarchical clustering implemented in HICLUS is often used as an alternative or as a supplementary technique to the basic model of MDS and takes the same form of data.

The matrix of (dis)similarities between a set of objects is used to define a set of non-overlapping clusters such that the more similar objects are joined together before less similar objects. The scheme consists of a series of clustering (levels). In the initial level each object forms a cluster, whilst at the highest level all the objects form a single cluster. In a hierarchical clustering scheme (HCS) there are exactly (p-1) levels where there are p objects.

The clustering scheme is hierarchical in the sense that once two objects have been joined together at a lower level of the scheme, they may not be split at a higher level.

Only the ordinal information in the data is used in defining the clusterings.

1.3 RELATION OF HICLUS TO OTHER PROGRAMS IN MDS(X)

HICLUS is commonly used as an interpretative aid in analysing configurations of points resulting from MDS analyses.

2. DESCRIPTION

2.1 DATA

HICLUS expects data in the form of a lower triangle matrix of (dis)similarity measures between a set of objects (stimuli). Any of the types of data suitable for input to MINISSA are suitable (q.v.).

It is often tempting to submit to HICLUS the solution distances from (say) a MINISSA run. This is not recommended since a MINISSA solution will be locally unstable in the following sense. The location of the stimulus points in the space is not uniquely defined, since each may be moved within a fixed region without affecting the goodness-of-fit. It is precisely the small distances affected by such movements which are crucial in the early stages of the HICLUS analysis. Users are therefore advised to submit the original data to HICLUS.

2.2 THE MODEL

A hierarchical clustering scheme (HCS) consists of a set of clusterings of a set of objects at increasing levels of generality. At the lowest level, each object is considered a separate cluster. At the next level the two most similar objects are merged to form a cluster. At each subsequent stage either the most similar individual objects remaining are joined together to form a new cluster or an object (or indeed cluster) is joined to the cluster to which it is most similar. At the highest level objects fall into one large, undifferentiated cluster.

2.2.0.1 A simple example

```
                    Objects:
                    C   B   E   D   F   A
Level:  0           .   .   .   .   .   .
        1           .   XXXX    .   .   .
        2           .   XXXX    .   XXXX
        3           XXXXXXXX    .   XXXX
        4           XXXXXXXX    XXXXXXXX
        5           XXXXXXXXXXXXXXXXXXXX
```

In this example, B and E are merged at level 1, F and A are merged at level 2, C is merged with the cluster (B,E) at level 3, D is merged with (F,A) at level 4, and finally (C,B,E) and (D,F,A) are merged into a single cluster at the fifth level.

Notice that once an object has been assigned to a cluster it may not "leave" that cluster. This is the defining characteristic of a hierarchical scheme.

The crucial question when defining a HCS is one which asks how we are to calculate the (dis)similarity between an object and an existing cluster.

Consider three objects, a, b and c. If b and c have been joined to form a cluster (b,c) then the question arises, how are we to find the dissimilarity of a to (b,c). We might take it to be equal to the dissimilarity between a and b or to that between a and c or some average of the two. Since we are committed to using only the ordinal information in the data we disregard the averaging approach and are left in the general case, where a cluster may consist of more than two objects, with two options.

2.2.0.2 The "minimum" method

Also known as the "connectedness" or "single-link" method, this approach defines the dissimilarity between a point and a cluster as the smallest of the dissimilarities between the external point and the constituent points in the cluster. This method tends to join single points to existing clusters and schemes resulting from it are often not easily amenable to substantive interpretation. The "level" value in this approach gives the length of the longest chain joining any two points in the cluster. The approach is chosen by specifying METHOD(1) on the PARAMETERS card.

2.2.0.3 The "maximum" method

Also known as the 'diameter' or 'complete link' method, this approach defines the dissimilarity between a point and a cluster to be the largest of the dissimilarities between it and the points constituting

the cluster. In this case the "level" gives the size of the diameter of the largest cluster at that level. This method is chosen by specifying METHOD(2) on the PARAMETERS card. The default option METHOD(3) allows for both methods to be used sequentially.

With perfect data, both methods will give rise to the same clustering. The degree of difference between them may be regarded as an indication of how far the data are not representable by the model.

2.2.1 The Algorithm

At each level:

1. The smallest dissimilarity (greatest similarity) coefficient in the data matrix is identified.

2. The row- and column-element corresponding to this coefficient are then merged to form a cluster (i.e. one row and one column are effectively removed from the matrix).

3. The (dis)similarity coefficients between the new cluster and each of the remaining elements (points or clusters) are calculated according to the METHOD chosen.

4. The matrix is reduced by one row and column and the program returns to step 1.

5. When all the points are thus merged the solution is output in the form of a histogram.

2.3 FURTHER OPTIONS

2.3.1 The SHADE parameter

The SHADE procedure was developed by Ling (1973) as a means of representing clustering information in matrix form. Highly similar points are placed close to each other and the actual matrix entry is replaced by a character whose darkness is proportional to the similarity between the row- and column-elements. This is actually achieved on the line printer by means of overprinting.

The program takes the original matrix and rearranges the rows and columns so that objects which fall in the same cluster are placed adjacent to each other. The dissimilarities are then split into fifteen groups, each of which is then represented by a particular overprinted character whose degree of darkness reflects the level of dissimilarity. The user is allowed choice in the definition of these groups.

2.3.1.1 When SHADE (1)

The specification of SHADE (1) on the PARAMETERS card instructs the program to take the rank order of (dis)similarities and divide these into fifteen groups.

2.3.1.2 When SHADE (2)

SHADE (2) by contrast takes the actual (dis)similarities and it is these which are divided into fifteen equal intervals. Obviously, this will not result in the same representation if the actual values are highly clustered.

2.3.1.3 When SHADE (3)

When this option is chosen, the user specifies a minimum and a maximum (dis)similarity value between which bounds the data will be split

into the fifteen intervals. The bounds are specified by means of the READ DATA control card (under its associated INPUT FORMAT card). Any values which are less than the minimum or greater than the maximum are automatically included in the highest and lowest interval respectively. This option is useful when there are a few outlying values which would otherwise distort the intervals.

2.3.1.4 When SHADE (4)

Alternatively, the user may specify, by using SHADE (4) each of the boundaries. In this case the READ DATA card will read in sixteen values which define the categories.

3. INPUT PARAMETERS

3.1 LIST OF PARAMETERS

Keyword	Default	Function
DATA TYPE	0	0: The data are similarities. 1: The data are dissimilarities.
METHODS	3	1: Only the minimum method is used. 2: Only the maximum method is used. 3: Both methods are used (independently).
SHADE	0	0: SHADE analysis is not performed. 1: Ranks of the distances are used. 2: Actual distances are used. 3: Bounds are specified. 4: Intervals are specified.

3.2 NOTES

1. The following cards are not valid with HICLUS.

   { # / N / NO } OF SUBJECTS

   DIMENSIONS

   ITERATIONS

   PLOT

   PUNCH

2. The { # / N / NO } OF STIMULI card may be replaced with { # / N / NO } OF POINTS

3. The input should be specified as floating-point (F-type) numbers and should be presented as a lower-triangle matrix without diagonal.

3.3 PROGRAM LIMITS

Maximum number of stimuli = 80

3.4 PRINT, PLOT AND PUNCH OPTIONS

3.4.1 PRINT options

Option	Description
HISTORY	A detailed history of the clustering is produced.

3.4.2 PLOT and PUNCH options

There are no plotting or punching options in HICLUS.

4. EXAMPLE

4.1 Example of a run

```
col 1                col 16
RUN NAME             HICLUS TEST DATA
N OF POINTS          10
INPUT FORMAT         (10F4.0)
PARAMETERS           DATA TYPE(1), METHODS(2)
READ MATRIX
  <data>
COMPUTE
FINISH
```

BIBLIOGRAPHY

Burt, R.S. (1976) Positions in Networks, Social Forces, 55,(1), 93-122.

Cermak, G.W. and P.C. Cornillo (1976) Multidimensional analyses of judgments about traffic noise, Journal of the Accoustical Society, 59, (6), 1412-20.

Desbarat, J.M. (1976) Semantic structure and perceived environment, Geographical Analysis, 8, (4), 453-467.

Everitt, B. (1974) Cluster Analysis, London: Heinemann.

Johnson, S.C. (n.d) A simple clustering statistic, Bell Laboratories, mimeo.

Johnston, J.N. (1976) Typology formation across socio-economic indicators, Sociological Economics, 10, (4), 167-171.

Ling, R.F. (1973) A computer generated aid for cluster analysis, Communications of the ACM 16, 355-61.

Perreault, W.D. and F.A. Russ (1976) Physical distribution service in industrial purchase decisions, Journal of Marketing, 40, (2), 3-10.

Preece, P.F.W. (1976) Mapping cognitive structure - Comparison of methods, Journal of Educational Psychology, 68,(1), 1-8.

Seligson, M.A. and J.A. Booth (1976) Political participation in Latin America - Agenda for research, Latin American Research, 11, (3), 95-119.

Shepard, R.N. (1974) Representation of structure in similarities data: problems and prospects, Psychometrika, 39, 373-421.

Ward, J.H. Jr. (1963) Hierarchical grouping to optimise an objective function, Journal of the American Statistical Association, 58, 236-244.

APPENDIX 1: RELATION OF HICLUS TO PROGRAMS NOT IN MDS(X)

For a full range of options regarding hierarchical and other clustering schemes, users are referred to the CLUSTAN package.

APPENDIX 2: TECHNICAL DESCRIPTION*

Let

$$\underset{\sim}{\Delta} \equiv \{\delta_{jk}\} \qquad j,k = 1,\ldots,p$$

be a matrix of similarity coefficients between objects j,k and

$$C_o, C_1, \ldots, C_{p-1}$$

be a sequence of clusterings each with an associated value α

$$\alpha_o, \ldots, \alpha_{p-1}$$

We assume that C_o is the weak clustering with $\alpha_o = 0$, that C_{p-1} will be the strong clustering and that $\alpha_{j-1} \leq \alpha_j$.

Johnson shows that if C_i is the lowest cluster at which objects j,k are joined then α_i defines the ultrametric distance between j,k.

He goes on to describe two methods of defining a HCS which are monotone invariant.

A2.1 The minimum method

1. C_o is the weak clustering with value α_o

2. Assume C_{i-1} with similarity δ defined on all pairs.
   Let δ_{jk} be the minimum entry $(\delta_{jk} \neq 0)$.
   Create C_i by merging j and k. $\alpha_i = \delta_{jk}$.

3. If j and k are clustered in C_i but not in C_{i-1} then the distance from the new cluster to any third object ℓ is

$$d((j,k),\ell) = \min\,(d(j,\ell), d(k,\ell))$$

---

*This appendix is based on Johnson (1967) which is used with permission.

4. The process is repeated until the strong clustering is obtained.

A2.2 The maximum method

The maximum method is identical except that at step 3

$$d((j,k),\ell) = \max\,(d(j,\ell),\ d(k,\ell))$$