***N*. BBDIAM (Branch-and-Bound clustering)**

Page No.

1. OVERVIEW

Concisely: BBDIAM (Branch and Bound clustering) partitions a single matrix of (dis)similarities into a specified number of clusters, identifying an upper bound by biased-sampling complete link cluster analysis, and then uses a branch-and-bound algorithm to minimize the 'partition diameter', that is, the maximum of the cluster 'diameters' (the maximum pairwise dissimilarity values within each group). (The same process is used in CONPAR, qv, when partitioning the matrix of dissimilarities between a set of subject matrices).

DATA:	2-way, 1-mode dis/similarities
TRANSFORMATION:	Nonmetric, invariant under monotone transformations
MODEL:	Minimizing partition diameter for a specified number of groups

ORIGIN

This version of branch-and-bound partitioning was developed by Michael Brusco and colleagues (Brusco 2003, Brusco and Stahl 2005). The algorithm is published in Brusco and Stahl and used within NewMDSX with their permission.

1.1 BRIEF DESCRIPTION OF BBDIAM

Clustering is often used as an alternative or as a supplementary technique to the basic model of MDS and generally takes the same form of data. BBDIAM uses an efficient new algorithm (Brusco 2003) to partition the input matrix into the number of clusters specified by the user. The process can be repeated in a single run for several different numbers of clusters. The content of the various partitions produced by BBDIAM may usefully be compared with the results of applying MINISSA to the same matrix. Given the number of ties that can occur in minimum diameter partitioning, it is likely that there are many alternative optima in large matrices. Brusco's algorithm is designed to reduce the chances that the process ends a subobtimal partition.

1.2 RELATION TO OTHER PROGRAMS IN NewMDSX

Like HICLUS (Johnson 1967) BBDIAM is a clustering model and not a spatial model. In contrast to HICLUS, the user of BBDIAM must pre-specify the number of partitions desired (and may specify several solutions)and the program carries out a separate partitioning process for each of the solutions. Unlike in HICLUS, objects which are placed together for one number of partitions will not necessarily be placed together in the next smaller number of partitions. The set of partitions produced is therefore a non-hierarchical clustering technique (Everitt 1974, Ward 1963)

The measure of the diameter of a partition which BBDIAM seeks to minimize is related to the Johnson's diameter measure (maximum, or complete link method) also used in HICLUS. An equivalent routine is used to partition the subjects' input matrices matrix of similarities in CONPAR.

Since the method does not develop an hierarchical system of clusters it is not relevant to display them as a dendrogram, as in HICLUS. The content of the various partitions produced by BBDIAM may, however, be mapped into a MINISSA or MRSCAL solution as a graphical aid to interpretation of the configuration.

2. DESCRIPTION OF THE PROGRAM

2.1 DATA

BBDIAM accepts as input either the lower triangle (without diagonal) or a full square symmetric data matrix. Each entry of this input matrix is a measure of (dis)similarity between the row-element and the column element. Commonly these are pair-wise ratings of similarity, but any symmetric measure may be used (including correlations, covariances if they are non-negative) and co-occurrences. Subject to the DATA TYPE specified, the matrix can be converted, if necessary, into dissimilarities.

2.2 THE ALGORITHM

The object of the algorithm is to develop a partition of a matrix of pairwise dissimilarity measures matrix into a series of clusters, each containing at least one item. Clusters are to be mutually exclusive, and exhaustive, in that all items are assigned to a cluster. A commonly-used criterion in clustering is to minimize the within-cluster sum of pairwise dissimilarities, but this has a tendency to produce clusters of approximately the same size, irrespective of the data. For this reason, the enhanced branch and bound algorithm employed by BBDIAM seeks instead to minimize the partition diameter, which is related to Johnson's diameter method for hierarchical clustering. The diameter of a cluster is the maximum pairwise dissimilarity index among objects in that cluster. The partition diameter is defined as the maximum of the cluster diameters. To minimize the diameter of the partition is to minimize the maximum dissimilarity index accross all subsets. An advantage of the partition diameter index is that it is not predisposed to produce clusters of particular sizes. It is also computationally simpler than minimizing the within-cluster sum of dissimilarities.

The efficiency of the branch-and-bound algorithm in minimizing the partition diameter depends on the quality of the upper bound. A good upper bound can frequently be established heuristically using a complete-link clustering algorithm, as in HICLUS Method(1). An algorithm for partitioning then applies two kinds of local-search operations - trial movement of each object from its current subset to each of the other subsets, and pairwise interchange of objects w.r.t. their subset memberships. These local-search operations are conducted until no further improvement is possible in the upper bound.

Because the resulting solutions may be sensitive to the initial partition, BBDIAM applies a probabilistic process that uses 100 replications and selects linkages using biased sampling. In particular, when considering the next linkage, the algorithm has a 50% chance of selecting the best linkage and a 50% chance of selecting the second best linkage. This biased-sampling version often produces better bounds than the deterministic version.

3. USING BBDIAM

BBDIAM expects data in the form of a lower triangle or a full symmetric matrix of (dis)similarity measures between a set of objects (stimuli). Any of the types of data suitable for input to MINISSA (and other 2W1M programs) are suitable. Note, however, that data values must be non-negative.

BBDIAM will produce the series of partitions defined by the CLUSTERS command, which may specify a list and/or a range of numbers of partitions, for example 2,5,8 or 2 TO 5 or 2, 4 TO 6. Specifying CLUSTERS to produce full enumeration of all possible combinations is not recommended for matrices with many more than 15 variables because of the time this may take to compute.

4. INPUT COMMANDS

Keyword	Function
N OF STIMULI [number]	Number of stimuli in the analysis

LABELS	[followed by a series of labels (<= 65 chars each on a separate line]	Optionally identify the stimuli. There should be as many labels as there are stimuli.
CLUSTERS	[number] [number list] [number] TO [number]	The number of clusters to be identified by partitioning the input matrix
READ MATRIX		Start reading the input matrix

4.1 PARAMETERS

All parameter keywords may be shortened to the first four letters. All subsequent mis-spellings are ignored.

Keyword	Default	Function
DATA TYPE	1	0: Lower-triangle matrix of similarities (high values mean high similarities between points). 1: Lower-triangle matrix of dissimilarities (high values mean high dissimilarities between points). 2: Full-symmetric matrix of similarities (high values mean high similarities between points). 3: Full-symmetric matrix of dissimilarities (high values mean high dissimilarities between points).

Note: N OF STIMULI may be replaced with N OF POINTS for BBDIAM.

Apart from PRINT DATA, there are no PRINT or PLOT options for BBDIAM.

PROGRAM LIMITS
Maximum no. stimuli = 80
Maximum no. of clusters = 80

4.2. OUTPUT

The output simply lists the content of each of the partitions requested. Giving names to the variables or stimuli when entering the matrix using the input Wizard, or use of the LABELS command in the input, considerably clarifies the cluster contents listed.

5. EXAMPLE

The following example (from Coxon 2007) partitions a matrix consisting of co-occurrence frequencies for a series of “drugs” elicited by a free-listing exercise by 68 subjects

```
RUN NAME       Coxon’s drugs example
N OF STIMULI   28
PARAMETERS     DATA TYPE(0)
CLUSTERS       3,5
LABELS         ALCOHOL
               AMPHETAMINE
               ASPIRIN
               BARBITURATES
```

```
              CAFFEINE
              CANNABIS
              CHOCOLATE
              COCAINE
              COUGH  MIXTURE
              CRACK
              ECSTASY
              GHB
              GLUE
              HEROIN
              IMMODIUM
              INSULIN
              KETAMINE
              LSD
              MAGIC-MUSHROOMS
              METHADONE
              PCP
              PENICILLIN
              POPPERS
              PROZAC
              STEROIDS
              TEMAZEPAM
              TOBACCO
              VIAGRA
READ MATRIX
   4
   6   4
   4  22  20
  49   2   9   3
  18  24   3  11   8
  47   2   8   2  58   8
   6  37   0  21   5  27   3
   9   4  50  19   9   6   9   0
   7  32   0  21   4  19   3  46   1
   5  45   0  18   6  25   3  38   1  38
   0  18   2  16   3   6   1  20   3  20  17
  13  14   3  13   7  23   6  23   6  24  20  11
   6  27   2  21   2  23   1  51   3  47  30  18  23
   6   4  36  14   6   3   6   0  35   0   0   9   4   2
   8   7  49  18   6   4   6   3  40   2   2   4   4   1  31
   2  30  11  24   1   9   1  23  10  24  22  26  10  20  12  11
   4  31   0  20   2  29   2  38   1  44  43  15  26  42   0   0  18
   9  25   3  15   5  39   4  19   3  21  31  12  18  23   4   0  18  32
   4  17  21  28   3   9   2  20  20  18  14  12  11  21  16  25  18  15   6
   2  24   6  30   2  12   1  23   7  25  23  29  15  22  11   4  35  23  20  15
   6   4  57  19   7   3   7   0  47   0   0   1   3   2  36  56   9   0   3  22   5
   3  28   3  21   3  15   1  26   5  25  30  28  14  21   8   5  30  22  21  13  34   3
   5   7  29  19   7   5   6   2  26   3   3   5   7   5  24  29   6   2   6  25   9  35
7
   3  22  21  21   6  11   4  11  18  10  13   9  10  14  11  19  17   6  14  21  11  24
13  25
   4  12  13  27   2   5   2  10  13  11  10  16  11  14  18  15  20  11  10  19  28  18
18  22  16
  55   4   8   3  52  13  50   5   8   4   3   2   7   3   6   8   2   2   7   3   2   6
3   4   7   3
   5   8  37  18   7   4   6   1  28   1   4   6   7   3  21  36  11   3   6  21   6  40
6  34  35  15   6
COMPUTE
FINISH
```

SOLUTION: The clusters requested are listed as follows:

```
OPTIMISED BRANCH-AND-BOUND CLUSTERING
                                        MINIMUM DIAMETERS
 PARTITIONS                   HEURISTIC SOLUTION     OPTIMAL SOLUTION

  3:                                 51.00                51.00
 -------------
```

```
  1) ALCOHOL, CAFFEINE, CANNABIS, CHOCOLATE, TOBACCO


  2) AMPHETAMINE, BARBITURATES, COCAINE, CRACK, ECSTASY, GHB, GLUE, HEROIN, KETAMINE,
LSD, MAGIC-MUSHROOMS, PCP, POPPERS, TEMAZEPAM


  3) ASPIRIN, COUGH  MIXTURE, IMMODIUM, INSULIN, METHADONE, PENICILLIN, PROZAC, STEROIDS,
VIAGRA
================================================================

OPTIMISED BRANCH-AND-BOUND CLUSTERING
                                      MINIMUM DIAMETERS
 PARTITIONS                   HEURISTIC SOLUTION     OPTIMAL SOLUTION
  5:                                 41.00                41.00


  1) ALCOHOL, CAFFEINE, CHOCOLATE, TOBACCO

  2) CANNABIS, ECSTASY, GLUE, HEROIN, LSD, MAGIC-MUSHROOMS

  3) IMMODIUM, PROZAC, TEMAZEPAM

  4) AMPHETAMINE, COCAINE, CRACK, GHB, KETAMINE, PCP, POPPERS

  5) ASPIRIN, BARBITURATES, COUGH  MIXTURE, INSULIN, METHADONE, PENICILLIN, STEROIDS,
VIAGRA
```

BIBLIOGRAPHY

Brusco, M.J. (2003) An enhanced branch-and-bound algorithm for a partitioning problem, British Journal of Mathematical and Statistical Psychology, 56, 83-92.

Brusco, M.J. & S.Stahl (2005) Branch-and-Bound Applications in Combinatorial Data Analysis, New York: Springer

Coxon, A.P.M. (2007) Using Multidimensional Scaling, London:Sage

Everitt, B S, S Landau & M. Leese. (2001) Cluster Analysis. Arnold

Johnson, S.C. (1967) Hierarchical clustering schemes, Psychometrika, 32, 241-254.

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

(APB , modified by APMC Jan 11 2006)