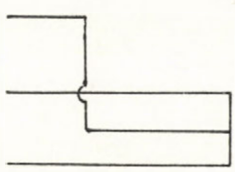
MTA Számitástechnikai és Automatizálási Kutató Intézete, Közlemények 24/1980.

Result of Cluster Analysis: Dendrogram
/An Algorithm for Displaying the Dendrogram/

Ferenc Juhász - Károly Tóth
Budapest

Many of the methods of cluster analysis consists in hierarchical procedures. The results of these procedures are contained most concisely in the so-called dendrogram.

The dendrogram is a binary tree the branches of which correspond to the mergers /separations/ of clusters. The dendrogram of several methods /Centroid, Median, Eigen.../ may have reversal i.e. it may contain the following isomorphic tree:



An important auxliary tool of the high-level software is the plotter. Displaying the dendrogram by means of the plotter is justified by two points of view:

(1) the dendrogram obtained can be published without any alteration,

(2) the reversal phenomenon is easy to handle.

It is difficult to get literary data on programs used for displaying dendrograms on a plotter [1, p.328], therefore it appeared desirable to develop the subroutines described below.

In order that the results obtained by different methods may be compared, the suitable version must be selected from the possible 2^{N-1} isomorphic variants with the use of some uniform principle. It is practical to choose such a relatively natural permutation of the elements that the dendrogram plotted on its order does not contain any intersection if not necessary.

The question arises what is the minimum information serving for describing the dendrogram.

(1) The hierarchical procedure generally constitues a series $(I(K), J(K))$, $K=1,\ldots,N-1$ which expresses that in step K the clusters I(K) and J(K) merge. /The algorithm identifies each cluster with the element of the lowest order occurring in it./

(2) It is natural that the clusters contain the elements ordered according to their serial number.

The above informations determine the permutation of the dendrogram elements univocally. This permutation is produced by the subroutine PERMUT. The dendrogram is displayed by the subroutine DENDPLOT called after this.

The subroutines pertaining to the concrete machine realization used in the program, i.e. the subroutines PLOTOPLEN, PLOTOLOS, MOVE, DRAW and PLOTVAL have been written for the computer CDC 3300 and they control the plotter CALCOMP 363. The first two serve for opening and closing the plotterfile, the other two for moving the pen, while the last one is a subroutine for writing characters.

Figures 1 and 2 show dendrograms concerning to the problem published in paper [2] /based on Bernstein's method/.

SUBROUTINE PERMUT (I,J,N,NL,IPERM,IL,JL,NEXT)

Formal parameters

I	integer array	(NL)	input:	in step K clusters I(K)
J	integer array	(NL)	input:	and J(K) merge
N	integer		input:	number of points
NL=N-1			input	
IPERM	integer array	(N)	output:	permutation of the serial numbers of the elements
IL	integer array	(NL)	workspace	
JL	integer array	(NL)	workspace	
NEXT	integer array	(N)	workspace	

SUBROUTINE DENDPLOT (C,H,I,J,IPERM,N,NL,S,PX,PY)

Formal parameters

C	real		input: horizontal dimension of the dendrogram (in 0.01 inches)
H	real		input: height of the characters (in 0.01 inches)
I	integer array	(NL)	input: as above
J	integer array	(NL)	input: as above
IPERM	integer array	(N)	input: as above
N	integer		input: as above
NL=N-1			input
S	real array	(NL)	input: levels belonging to the mergers
PX	real array	(N)	workspace
PY	real array	(N)	workspace

External subroutines

PLOTOPEN (IDSI,IBLK)

IDSI	integer	input: identification of the plotterfile
IBLK	integer	input: maximum number of blocks

PLOTCLOS

MOVE (X,Y)

X	real	input: } drawing pen in lifted position
Y	real	input: } is moving into point (X,Y)

DRAW (X,Y)

X	real	input: } pen is drawing a line from the
Y	real	input: } actual point into point (X,Y)

PLOTVAL (IFORMAT,IVALUE,NCHAR,HEIGHT,ANGLE)

IFORMAT	integer array	input: format required for drawing the value of IVALUE
IVALUE	integer	input: value to be drawn
NCHAR	integer	input: number of characters to be drawn
HEIGHT	real	input: height of numbers (in 0.01 inches)
ANGLE	real	input: rotation angle of characters (in radians)

```
      SUBROUTINE PERMUT(I,J,N,NL,IPERM,IL,JL,NEXT)
C
C        CONSTRUCT THE PERMUTATION OF THE NODES
C
      DIMENSION I(NL),J(NL),IPERM(N)
      DIMENSION IL(NL),JL(NL),NEXT(N)
C
C        INITIALIZE IL, JL, NEXT
C
      DO 1 K=1,N
      NEXT(K)=0
    1 CONTINUE
      DO 2 K=1,NL
      IK=I(K)
      JK=J(K)
      IL(K)=NEXT(IK)
      JL(K)=NEXT(JK)
      NEXT(IK)=K
      NEXT(JK)=K
    2 CONTINUE
      DO 3 K=1,NL
      NEXT(K)=N
    3 CONTINUE
      DO 4 K=1,NL
      LK=IL(K)
      IF(LK.NE.0) NEXT(LK)=K
      LK=JL(K)
      IF(LK.NE.0) NEXT(LK)=K
    4 CONTINUE
C
C        GO OVER THE TREE
C
      DO 5 IK=1,NL
      IF(I(IK).NE.1) GO TO 5
      K=IK
      GO TO 6
    5 CONTINUE
    6 IST=0
   11 IF(IL(K).NE.0) GO TO 12
      IST=IST+1
      IPERM(IST)=I(K)
      IL(K)=-1
   12 IF(IL(K).EQ.-1) GO TO 21
      KN=IL(K)
      IL(K)=-1
      K=KN
      GO TO 11
   21 IF(JL(K).NE.0) GO TO 22
      IST=IST+1
      IPERM(IST)=J(K)
      JL(K)=-1
   22 IF(JL(K).EQ.-1) GO TO 31
      KN=JL(K)
      JL(K)=-1
      K=KN
      GO TO 11
   31 K=NEXT(K)
      IF(K.EQ.N) GO TO 41
      GO TO 11
   41 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE DENDPLOT(C,H,I,J,IPERM,N,NL,S,PX,PY)
C
C        KNOWING THE PERMUTATION OF THE NODES DISPLAY THE DENDROGRAM
C
      DIMENSION I(NL),J(NL),IPERM(N),S(NL)
      DIMENSION PX(N),PY(N)
      DIMENSION IFORMAT(1)
C
C        INITIALIZE THE PARAMETERS
C
      IFORMAT(1)=4H(I3)
      OX=90.
      OY=90.
      H67=H*6./7.
      H97=H*9./7.
      PI=3.1415926
      PI2=PI/2.
      SMAX=0.
      DO 4 K=1,NL
      IF(S(K).GT.SMAX) SMAX=S(K)
    4 CONTINUE
      CALL PLOTOPEN(11,500)
C
C        DISPLAY THE CHARACTERS
C
      X=OX
      Y=OY-3.*H67
      DO 1 K=1,N
      X=X+H97
      CALL MOVE(X,Y)
      L=IPERM(K)
      CALL PLOTVAL(IFORMAT,L,3,H,PI2)
    1 CONTINUE
C
C        DRAW THE TREE
C
      DO 2 K=1,N
      L=IPERM(K)
      PX(L)=OX+H67+H97*(K-1)
      PY(L)=OY
    2 CONTINUE
      DO 3 K=1,NL
      IK=I(K)
      JK=J(K)
      BX=PX(IK)
      BY=PY(IK)
      SY=OY+C*S(K)/SMAX
      EX=PX(JK)
      EY=PY(JK)
      CALL MOVE(BX,BY)
      CALL DRAW(BX,SY)
      CALL DRAW(EX,SY)
      CALL DRAW(EX,EY)
      PX(IK)=(BX+EX)/2.
      PY(IK)=SY
    3 CONTINUE
      CALL PLOTCLOS
      RETURN
      END
```

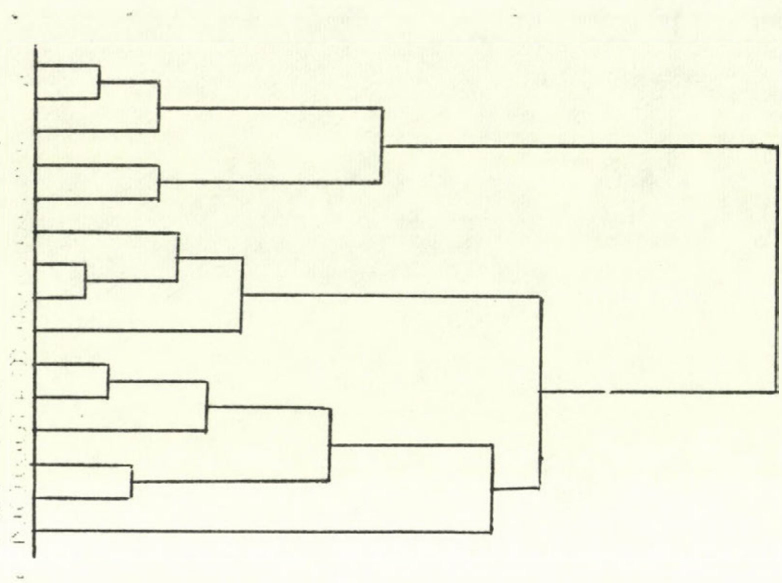


Figure 1 Furthest Neighbour

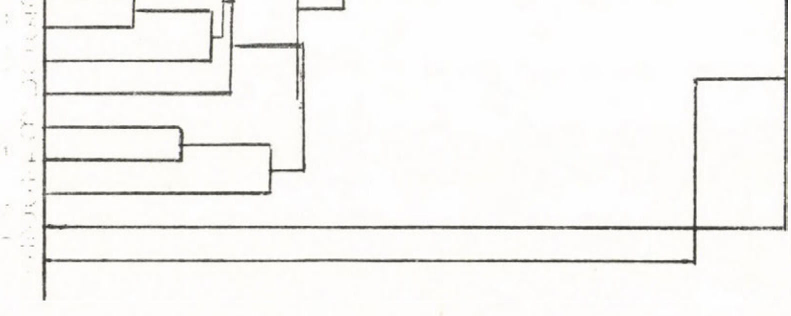


Figure 2 Median

REFERENCES

[1] Anderberg,M.R. /1973/: Cluster Analysis for Applications. Academic Press, New York - San Francisco - London.

[2] Juhász,F. /1978/79/: The Analysis of the Distribution of Blood Groups by Means of Clustering Methods. Mitt.Arch.Inst. 8/9.

ÖSSZEFOGLALÁS

Juhász F. - Tóth K.

A dolgozat algoritmust közöl dendogram plotteren való megjelenitésére.

Результат кластерного анализа: дендорограмма

Юхас Ф. - Тот К.

В работе дается алгоритм для изображения дендрограммы на плоттере.