枝状图法的两个电算程序

范守志
（中国科学院海洋研究所）

枝状图（dendrograph）法由于麦卡蒙（R.B.McCamm）的发展，在数学生态学和统计地质学的研究中得到了广泛的应用。它可用于Q型分析和R型分析，计算结果可绘制成二维枝状图，一目了然地显示出站位间或种属、变量间的群分关系。方法的实施多在电子计算机上进行。

已发表的源程序有三种，系用BCY-乙语言（用于109-乙机）、DJS-20语言及Fortran语言编制。

现提供两份用DJS-6的ALGOL语言编制的源程序，适用于108-乙机。这两份是做Q型分析用的，若做R型分析可仿之改写。其中，程序A以$\cos\theta$相似性系数作为站位间的相似性指标，程序B以C_J值即Jaccard系数作为相似性指标。

两份程序都是严格地依逐步聚类的要求进行的，即计算、判别、挑选、打印输出、合群，然后再循环上述步骤，直至所有站位都合成一群为止。在计算中的任何阶段，每群的代号均取为该群所含各站中站号（统计编号）最小者。每循环一次，都打印输出五个数据：

所合两群中“旧群”（编号较小的群）的代号；

所合两群中“新群”（编号较大的群）的代号；

这两群间的相似性指标值；

“旧群”中所含的站位个数；

“新群”中所含的站位个数。

这后两个数据便于绘制枝状图时留好空间，可一次成图。

显然，最后一次循环的输出中，第一个数据必然是1，最后两个数据之和等于参加统计的站位总数。

作图时，应由最后一次循环的输出开始，逆序而行。

因108-乙机内存较小，而输入数据阵及作为中间结果的相似系数对角阵往往很大，故这两个阵均不保留，每轮合群直接在数据阵中进行。这就节省了内存，但相应地增大了计算量。并且程序A，B需要分立。好在地质研究中不用C_J系数，而生物研究中目前大多用C_J系数。

一、程　序A

说明

N——站位总数；M——每站中参数的个数；X——二维数组$(X_{ik})_{N\times M}$，其中X_{ik}是第i个站位中第k种参数的测定值，既可用绝对含量也可用百分含量，但要统一；Y和Z——均为一维数组，Y_i是第i群的编号，Z_i是第i群中含有的站位个数（起初，$Y_i=i$而$Z_i=1$）；Q——两群间的$\cos\theta$值。

第i群和第j群间的Q值依下式计算：

$$Q=\cos\theta=\frac{\sum_{K=1}^{M}X_{ik}X_{jk}}{\sqrt{\sum_{K=1}^{M}X_{ik}^2}\sqrt{\sum_{K=1}^{M}X_{jk}^2}}$$

每次合群时应使原始各站的数据在平均时有相等的权重。因此当A、B两群（A＜B）合为新的A群时，取

$$X_{AK}=\frac{Z_AX_{AK}+Z_BX_{BK}}{Z_A+Z_B},$$

$$(K=1,2,\cdots\cdots M)$$

源程序A

```
'BEGIN'
'INTEGER'N,M;READI(N,M);
'BEGIN'
```

```
    'INTEGER'I,J,K,A,B,R;
    'REAL'U,V,W,Q,QMAX;
    'INTEGER''ARRAY'Y〔1:N〕,
Z〔1:N〕;
    'ARRAY'X〔1:N,1:M〕;
    'SWITCH'SW:=L;
    R:=N; INPUT(X);
    'FOR'K:=1'STEP'1'UNTIL'N'DO'
    'BEGIN'
    Y〔K〕:=K; Z〔K〕:=1
    'END';
L:A:=B:=0; QMAX:=0;
    'FOR'I:=2'STEP'1'UNTIL'R'DO'
    'FOR'J:=1'STEP'1'UNTIL' I-1
'DO'
    'BEGIN'
    U:=V:=W:=0;
    'FOR'K:=1'STEP'1'UNTIL'M
'DO'
    'BEGIN'
    U:=U+X〔I,K〕×X〔I,K〕;
    V:=V+X〔J,K〕×X〔J,K〕;
    W:=W+X〔I,K〕×X〔J,K〕
    'END';
    U:=SQRT(U); V:=SQRT(V);
    U:=U×V;
    'IF'U=0'THEN'Q:=0'ELSE'
    Q:=W/U;
    'IF'Q'GR'QMAX'THEN'
    'BEGIN'
    QMAX:=Q; A:=J;B:=I
    'END'
    'END';
    DUMMY(5); OUTPUTI(Y〔A〕,Y
〔B〕);
    OUTPUTR(QMAX); OUTPUTI(Z
〔A〕,Z〔B〕);
    'FOR'K:=1'STEP'1'UNTIL'M'DO'
    'BEGIN'
    X〔A,K〕:=(Z〔A〕×X〔A,K〕+
Z〔B〕×X〔B,K〕)/(Z〔A〕+Z〔B〕);
    'FOR'I:=B'STEP'1'UNTIL'R-1
'DO'
    X〔I,K〕:=X〔I+1,K〕
    'END';
    Z〔A〕:=Z〔A〕+Z〔B〕;
    'FOR'K:=B 'STEP'1'UNTIL' R-1
'DO'
    'BEGIN'
    Y〔K〕:=Y〔K+1〕; Z〔K〕:=Z〔K+1〕
    'END';
    R:=R-1;
    'IF'R'GQ'2'THEN''GOTO'L
    'END'
    'END'
    XXX
```

二、程　序 B

说明

程序中的N、M、X、Y和Z的涵义同于程序A。变量CJ指的是两群间的C_J系数，即

$$C_J=\frac{\text{在两群中同时出现的种数}}{\text{两群总共涉及的种数}}$$

$$=\frac{W}{U+V-W},$$

这里W是共有种的种数，U和V分别为在两群中出现的种数。

数组X是整型的，它的元素X_{ik}是第i站处第K种的个体数，只能是零或正整数。本程序也允许直接以有—无型数据输入，即X_{ik}只取0或1。

源程序B

```
    'BEGIN'
    'INTEGER'N,M; READI(N,M);
    'BEGIN'
    'INTEGER'I,J,K,A,B,R,U,V,W;
    'REAL'CJ,CJMAX;
    'INTEGER''ARRAY'X〔1:N,1:M〕,
    Y〔1:N〕, Z〔1:N〕;
```

```
    'SWITCH'SW:=L;
    R:=N; INPUTI(X);
    'FOR'I:=1'STEP'1'UNTIL'N'DO'
    'BEGIN' Y〔I〕:=I; Z〔I〕:=1;
    'FOR'K:=1'STEP'1'UNTIL'M'DO'
    'IF'X〔I,K〕'NQ'0'THEN'X〔I,K〕:=1
    'END';
L:  A:=B:=0; CJMAX:=0,
    'FOR'I:=2'STEP'1'UNTIL'R'DO'
    'FOR'J:=1'STEP'1'UNTIL'I-1
'DO'
    'BEGIN' U:=V:=W:=0;
    'FOR'K:=1'STEP'1'UNTIL'M
'DO'
    'BEGIN' U:=U+X〔I,K〕;
    V:=V+X〔J,K〕;
    W:=W+X〔J,K〕×X〔I,K〕
    'END';
    U:=U+V-W;
    'IF'U=0'THEN'CJ:=0'ELSE'
    CJ:=W/U;
    'IF'CJ'GR'CJMAX'THEN''BEGIN'
    CJMAX:=CJ; A:=J; B:=I'END'
    'END';
    DUMMY(5); OUTPUTI(Y〔A〕,Y
〔B〕);
    OUTPUTR(CJMAX);
    OUTPUTI(Z〔A〕,Z〔B〕);
    'FOR'K:=1'STEP'1'UNTIL'M'DO'
    'BEGIN'
    'IF'X〔A,K〕=0'THEN'X〔A,K〕:=X
〔B,K〕;
    'FOR'I:=B'STEP'1'UNTIL'R-1
'DO'
    X〔I,K〕:=X〔I+1,K〕
    'END';
    Z〔A〕:=Z〔A〕+Z〔B〕;
    'FOR'K:=B'STEP'1'UNTIL'R-
1'DO'
    'BEGIN'Y〔K〕:=Y〔K+1〕;
    Z〔K〕:=Z〔K+1〕
    'END';
    R:=R-1;
    'IF'R'GR'2'THEN''GOTO'L
    'END'
    'END'
    XXX
```

三、注 意 事 项

（一）数据纸带穿孔格式为1.5米左右的空白段（头段）号码键，一个正整数（站位总数）分号；逗号，一个正整数（参量个数）分号；XXX（停机号），1.5米左右的空白段的第一站的M个数据（每个后有分号；），小段空白；第二站的M个数据（每个后有；），小段空白；……第N站的M个数据（每个后有；），XXX（停机号），1.5米左右空白（尾段）。

（二）空白站位（它的M个参量值全为0）可以参加运算，程序将认为它与任何站位间的Q值或C_J值为零。

（三）由于C_J系数本身的缺欠，程序B不如A。因为枝状图法要求各次循环合群时挑出的C_J值较上一循环的C_J小。但至少在理论上说来，C_J系数不能保证这点。而Q因子无此弊病。

参 考 文 献

〔1〕中国科学院地质研究所，1978。数学地质引论。地质出版社。第十一章。

〔2〕於崇文等，1980。数学地质的方法与应用。冶金工业出版社。第三篇。

〔3〕Mather, P.M., 1976. Computational methods of multivariate Analysis in Physical Geography. John Wiely & Sons. p.308—399.

〔4〕McCamm, R.B., 1968. Geol. Soc. Amer. Bull. 79(11):1663—1670.