 Kineziologija, Vol. 15 (1983.) str. 5—8

KONSTANTIN MOMIROVIĆ, KSENIJA BOSNAR
JANEZ ŠTALEC i FRANJO PROT
Fakultet za fizičku kulturu Sveučilišta u Zagrebu

Primljeno 12. 11. 1982.

HERAKLIT — ALGORITAM I PROGRAM ZA METRIČKO MULTIDIMENZIONALNO SKALIRANJE OBJEKATA OPISANIH NAD SKUPOM NOMINALNIH VARIJABLI*

SAŽETAK

Algoritam HERAKLIT i njemu pridruženi program izvodi metričko multidimenzionalno skaliranje skupa objekata opisanih nad skupom nominalnih varijabli transformiranih u binarni oblik. Mjere sličnosti algoritam definira kao normirane skalarne produkte vektora objekata. Dimenzije algoritam definira projekcijama objekata na svojstvene vektore matrice mjera sličnosti, normirane na svojstvene vrijednosti, a njihov broj određuje tako da se može reproducirati 90% traga matrice mjera sličnosti. Ove dimenzije algoritam zatim transformira u kosi parsimonijski sustav ortonormalnim transformacijama. Za obje solucije algoritam određuje relacije između varijabli i latentnih dimenzija; te su relacije definirane kao normirani skalarni produkti između vektora varijabli i vektora latentnih dimenzija. Za kosu soluciju algoritam određuje i koordinate vektora varijabli u sustavu koji je definiran latentnim dimenzijama.

1. ALGORITAM

U eksperimentalnim nacrtima u psihologiji i ostalim antropologijskim znanostima objekti istraživanja veoma su često opisani na nominalnim mjernim skalama. Nekvantitativni podaci prikupljeni u takvim istraživanjima mogu se efikasno tretirati metričkim postupcima ukoliko se nominalne varijable prethodno binariziraju.

Predložen je algoritam i napisan program koji analizira latentnu strukturu skupa objekata, opisanih nad skupom nominalnih varijabli transformiranih u binarni oblik, tehnikom metričkog multidimenzoinalnog skaliranja.

1.1 Prethodne opreacije

Neka je

$$N = \{N_j;\ j=1,\dots,m\}$$

skup nominalnih varijabli N_j koje sadrže l_j kategorija K_{jp}, $p=1,\dots,l$ tako da je $q = \sum_{j=1}^{m} l_j$ ukupan broj kategorija u skupu N, i neka za kategorije $K_{jp} \varepsilon N_j$ vrijedi $K_{ip} \cap K_{jp'} = 0$ ako $P \neq p'$. Neka je $E = \{e_i;\ i=1,\dots,n\}$ skup objekata opisanih nad skupom N. Neka je svaki $e_i \varepsilon E$, opisan varijablama $N_j \varepsilon N$, reprezentiran vektorom.

$$B^T_{ij} = (b_{ij},\dots,b_{ijp},\dots,b_{ijl}\rangle \dots \qquad i=1,\dots,n \quad j=1,\dots,m$$

tako da je

$$b_{ijp} = 1 \text{ ako } e_i \varepsilon K_{jp}, \text{ a} \qquad i=1,\dots,n$$

$$b_{ijp} = 0 \text{ ako } e_i \varepsilon K_{jp}, \qquad j=1,\dots,m \quad p_j=1,\dots,l_j$$

gdje je $K_{jp} \cap N_j$ podskup definiran kategorijom p nominalne varijable N_j.

Neka su vektori B_j organizirani u matrice

$$B_j = (B^T_{ij}), \qquad j=1,\dots,m$$

a matrice B_j u supermatricu

$$B = (B_j).$$

Očito, B potpuno opisuje skup objekata E nad skupom nominalnih varijabli N.

1.2 Mjere sličnosti

Skalarni produkti vektora objekata

$$C = B\,B^T = (c_{ik}) \qquad i=1,\dots,n \quad k=1,\dots,n$$

formiraju Hemingove mjere sličnosti objekata iz E opisanih nad skupom N. Očito, u svakom elementu c_{jk}, $j \neq k$ matrice C je broj obilježja iz N, zajedničkih paru objekata $\{e_i, e_j\}$; naravno, tr $C_{ij} = m \ \forall\ i$, zbog mutualne ekskluzivnosti kategorija $K_{jp} \ \varepsilon\ N_j$.

Operacijom normiranja

$$(\text{diag } C)^{-1/2} C\ (\text{diag } C)^{-1/2} = Q$$

definira se Q, matrica normiranih mjera sličnosti objekata $e_i \varepsilon E$, sa svojstvom da je njen trag jednak ukupnom broju objekata (tr $Q = n$).

1.3 Ortogonalna solucija

Neka su λ_r; $r=1,\dots,s$ $s = \min(n, m-q+1)$ nenulte svojstvene vrijednosti matrice Q, i neka su X_r; $r=1,\dots,s$ njima pridruženi svojstveni vektori, normireni tako da je $X^T_r X_r = I \ \forall\ r$.

---

* Izloženo na VII kongresu psihologa SFRJ.

Neka se razloži matrica mjera sličnosti Q na dvije aditivne matrice.

$$Q = \sum_{r=1}^{t} X_r \lambda_r X_r^T + \sum_{r=t+1}^{s} X_r \lambda_r X_r^T,$$

gdje je t određen tako da suma zadržanih nenultih svojstvenih vrijednosti $\sum_{r=1}^{t} \lambda_r$ iscrpljuje 90% traga matrice Q.

Neka je $\Lambda = (\lambda_r)$; $r=1,\ldots,t$ dijagonalna matrica svojstvenih vrijednosti, uređenih tako da je $\lambda_1 > \lambda_2 > \ldots > \lambda_r$, a $X = (X_r)$; $r = 1,\ldots,t$ matrica njima pridruženih svojstvenih vektora.

U matrici

$$H = X \Lambda^{1/2}$$

bit će ortogonalne projekcije vektora objekata na denormirane lijeve svojstvene vektore matrice B.

Naravno, H je faktorska matrica matrice Q, jer je

$$L = H H^T = \sum_{p=1}^{t} \Lambda_p X_p X_p^T = Q - \sum_{p=t+1}^{s} \Lambda_p X_p X_p^T$$

reprodukcija matrice mjera sličnosti na temelju prvih t lijevih svojstvenih vektora matrice B.

1.4 Neortogonalna solucija

Transformacijom

$$H\,T = K = (k_{ir}) \qquad i=1,\ldots,n \quad r=1,\ldots,t$$

koja maksimizira funkciju jednostavnosti

$$W = u \sum_{i=1}^{n} \sum_{r=1}^{t} k_{ir}^4 - \sum_{r=1}^{t} \left(\sum_{i=1}^{n} k_{ir}^2\right)^2$$

uz uvjet $T^T T = T T^T = I$, gdje je, očito, T matrica relacija ortogonalnog i neortogonalnog koordinatnog sustava, moguće je formirati taksonomski koordinatni sustav za objekte iz E opisane na skupom N.

U matrici

$$\Psi = T^T \Lambda T$$

su relacije između kosih dimenzija, a u matrici

$$A = K\Psi^{-1}$$

koordinate objekata u kosom koordinatnom sustavu.

Matrica kosinusa kuteva između neortogonalnih dimenzija je

$$M = (\mathrm{diag}\ \Psi^{-1/2}\ \Psi\ (\mathrm{diag}\ \Psi)^{-1/2}.$$

Relacije između vektora varijabli i vektora kosih dimenzija su u matrici mjera kongruencije

$$G = (\mathrm{diag}\ B^T B)^{-1/2}\ B^T K\ (\mathrm{diag}\ K^T K)^{-1/2},$$

a koordinate vektora varijabli u sustavu kosih dimenzija su u matrici sklopa

$$\bar{S} = G\,M^{-1}.$$

2. PROGRAM

HERAKLIT, program za metričko multidimenzionalno skaliranje objekata opisanih nad skupom nominalnih varijabli transformiranih u binarni oblik, napisan je u verziji 5.1 meta jezika SS (STATISTICAL SYSTEM, Zakrajšek, Štalec i Momirović, 1974). Program pretpostavlja da je skup nominalnih varijabli prethodno transformiran u binarni oblik, da ni broj binarnih varijabli ni broj objekata ne prelazi 250, i da su programu pridružene naredbe za opis podataka koje zahtijeva SS operacija INPUT.

```
OUTPUT (DEVICE=PRI)
*          *** H E R A K L I T ***
*
*          RĐAVI SU SVJEDOCI OČI I UŠI LJUDIMA
*          KOJI IMAJU BARBARSKE DUŠE.
*
*                              HERAKLIT IZ EFESA.
*
* OVAJ JE PROGRAM NAMIJENJEN ONIMA
* KOJI IMAJU POTREBU DA DOKAZUJU ONO ŠTO JE OČITO.
*
* PROGRAM HERAKLIT IZVODI METRIČKO MULTIDIMENZIONALNO
* SKALIRANJE SKUPA OBJEKATA OPISANIH NAD SKUPOM
* NOMINALNIH VARIJABLI TRANSFORMIRANIH U BINARNI
* OBLIK. MJERE SLIČNOSTI DEFINIRANE SU KAO NORMIRANI
* SKALARNI PRODUKT VEKTORA OBJEKATA. DIMENZIJE SU
* DEFINIRANE SVOJSTVENIM VEKTORIMA MATRICE MJERA
* SLIČNOSTI, NORMIRANIM NA SVOJSTVENE VRIJEDNOSTI, A
* NJIHOV BROJ ODREĐEN TAKO DA SE MOŽE REPRODUCIRATI
* 90 POSTO TRAGA MATRICE MJERA SLIČNOSTI. OVE DIMENZIJE
* PROGRAM ZATIM TRANSFORMIRA U KOSI PARSIMONIJSKI
```

```
* SUSTAV ORTONORMALNIM TRANSFORMACIJAMA. ZA OBJE
* SOLUCIJE PROGRAM IZRAČUNAVA RELACIJE IZMEĐU VARIJABLI
* I LATENTNIH DIMENZIJA DEFINIRANE KAO NORMIRANI
* SKALARNI PRODUKTI IZMEĐU VEKTORA VARIJABLI I VEKTORA
* LATENTNIH DIMENZIJA, A ZA KOSU PARSIMONIJSKU SOLUCIJU
* I KOORDINATE VEKTORA VARIJABLI U SUSTAVU LATENTNIH
* DIMENZIJA
*
* PROGRAM PRETPOSTAVLJA DA SU NOMINALNE VARIJABLE
* VEĆ BINARIZIRANE, I DA NI BROJ VARIJABLI NI
* BROJ OBJEKATA NE PREMAŠUJE 250.
*
* PROGRAMU TREBA PRIDRUŽITI JEDNU SEQUENCE I ONOLIKO
* VARIJABLE NAREDBI KOLIKO JE BINARNIH VARIJABLI
* A ZATIM BLOK SA PODACIMA.
*
  TEXT(TEXT =            HERAKLIT).
  HEADING (TEXT = SKALIRANJE OBJEKATA OPISANIH NOMINALNIM VARIJABLAMA)
*
*          BLOK 1. MJERE SLIČNOSTI
*
* TEXT (TEXT =              MJERE SLIČNOSTI)
* INPUT (SCORE = B)
* MULT (A=B, B=B, TB, M=QR)
* SCALE (C=QR, R=Q)
* PRINT MATRIX=Q, TEXT=MATRICA MJERA SLIČNOSTI)
*
*          BLOK 2. ORTOGONALNA SOLUCIJA.
*
  TEXT (TEXT = ORTOGONALNA SOLUCIJA)
  DIAGONALISATION (R=Q, NAMEX=NMX)
  HOTELLING (TRACE=0.900. NAMEF=NMF, F=H)
  PRINT (MATRIX=H, T, TEXT=ORTOGONALNA SOLUCIJA)
  TRANSPOSE (OLD=H, NEW=HT)
  MULT (A=H, B=H, TB, M=L)
  STATISTICS (SCORE=HT, S, Z=ZH)
  DELETE (MATRIX=HT)
  DELETE (MATRIX=ZH)
*
*          BLOK 3. NEORTOGONALNA SOLUCIJA.
*
  TEXT (TEXT=NEORTO GONALNA SOLUCIJA)
  VARIMAX (F=H, TAU=TT, FN=KT, NAMEF=NMK)
  PRINT (MATRIX=TT, T, TEXT=TRANSFORMACIJSKA MATRICA)
  PRINT (MATRIX=KT, T, TEXT=NEORTOGONALNA SOLUCIJA)
  MULT (A=TT, B=L, M=TTL)
  MULT (A=TTL, B=TT, TB M=C)
  PRINT (MATRIX=C, TEXT=RELACIJE DIMENZIJA)
  INVERSION (R=C, RINV=CINV)
  MULT (A=KT, TA, B=CINV, M=A)
  PRINT (MATRIX=A, TEXT=KOORDINATE OBJEKATA)
  SCALE (C=C, R=M)
  PRINT (MATRIX=M, TEXT=KOSINUSI KUTEVA DIMENZIJA)
  TRANSPOSE (OLD=KT, NEW=K)
  STATISTICS (SCORE=K, S, Z=ZK)
  DELETE (MATRIX=K)
  DELETE (MATRIX=ZK)
  MULT (A=KT, B=KT, TB, M=KK)
  MULT (A=KT, B=B, M=CON)
  MULT (A=B, TA, B=B, M=BUB)
DIAGMULT (A=CON, T, D=BUB, C= —0,5, L, M=DBCON)
DIAGMULT (A=DBCON, D=KK, C= —0,5, R, M=G)
```

```
PRINT (MATRIX=G, TEXT=RELACIJE VARIJABLI I   DIMENZIJA)
INVERSION (R=M, RINV=MINV)
 MULT (A=G, B=MINV, M=S)
PRINT (MATRIX=S, TEXT=SKLOP VARIJABLI)
*
*          KRAJ PROGRAMA HERAKLIT
*
```

3. LITERATURA

1. Cattell, R. B: Handbook of multivariate experimental psychology; Rand McNally, Chicago, 1966.
2. Fulgosi, A.: Faktorska analiza; Školska knjiga, Zagreb, 1979.
3. Kruskal, J. B.: Multidimensional scaling by optimizing goodness to fit to a nonmetric hypothesis; Psychometrika, 29, 1—27, (1964A).
4. Kruskal, J. B.: Nonmetric multidimensional scaling: A numerical method. Psychometrika, 29, 115—129, (1964B).
5. Kruskal, J. B., M. Wish: Multidimensional scaling. Sage University paper, Beverly Hills and London, 1979.
6. Momirović, K., L. Szirovicza, M. Gredelj and V. Dobrić: Cattell — Algoritam and programme for the determination of polar taxone based on non-quantitive data; Collegium Antropologicum, 4, 41—44, 1980.
7. Štalec, J., K. Momirović i E. Zakrajšek: SS — Statistical system. Fakultet za fizičku kulturu Sveučilišta u Zagrebu i Sveučilišni računski centar, Zagreb, 1981.
8. Torgerson, W. S.: Theory and methods of scaling. John Wiley, New York, 1958.
9. Torgerson, W. S.: Multidimensional scaling of similarity. Psychometrika, 30, 379—393, 1965.
10. Zakrajšek, E., J. Štalec i K. Momirović: SS — programski sistem za multivarijantnu analizu podataka. Zbornik I Simpozija »Kompjuter na Sveučilištu«, 68, 1—16, 1974.

Momirović, K.; Bosnar, K.; Stalec, J.; Prot, F.

UDC 519.26 : 518.5

HERAKLIT — AN ALGORITHM AND PROGRAM FOR METRIC MULTIDIMENSIONAL SCALING OF OBJECTS DESCRIBED ABOVE A SET OF NOMINAL VARIABLES

multidimensional scaling / nominal variables

The HERAKLIT algorithm and its associated program perform metric multidimensional scaling of a set of objects described above a set of nominal variables transformed into the binary form. Measures of similarity are defined by the algorithm as normalized scalar products of objects vectors. The algorithm defines dimensions by projects onto eigenvectors of the similarity measures matrix. The projections are normalized to eigenvalues. The number of dimensions is determined by the algorithm in such a way that 90% of the trace of the similarity measures matrix can be reproduced. These dimensions are then transformed by the algorithm into an oblique parsimonial sy-stem by means of orthonormal transformations. The algorithm determines relations between variables and latent dimensions for both solutions; these relations are defined as normalized scalar products of vectors of the variables and vectors of latent dimensions. For the oblique solution the algorithm also determines vector coordinates of variables in the system defined by latent dimensions.

Константин Момирович, Ксения Боснар, Янез Шталец, Франье Прот
Факультет физической культуры, Загреб

ГЕРАКЛИТ — АЛГОРИТМ И ПРОГРАММА ДЛЯ МЕТРИЧЕСКОГО МУЛЬТИИЗМЕРИТЕЛЬНОГО ШКАЛИРОВАНИЯ ОБЬЕКТОВ, ОПИСАННЫХ НАД МНОЖЕСТВОМ НОМИНАЛЬНЫХ ПЕРЕМЕННЫХ

Алгоритм ГЕРАКЛИТ и соответствующая программа производят мультиизмерительное шкалирование множеств объектов, описанных над множеством номинальных переменных, которые трансформированы в бинарную форму. Измерения сходства определяются при помощи алгоритма как нормированные продукты векторов объектов. Алгоритм орределяет измерения при помощи проекций объектов на характерные величины, а их число определяется таким способом, чтобы обеспечить возможность воспроизведения 90% следов матрицы измерений сходства. Алгоритм потом трансформирует эти измерения в косую парсимоническую систему при помощи ортонормальных трансформаций. Для того и другого решения алгоритм определяет соотношения между переменными и латентными измерениями; эти соотношения определяются как нормальные шкаларные продукты между векторами переменных и векторами латентных измерений. Для косого решения алгоритм определяет также и координаты векторов переменных в системе, которая определяется при помощи латентных измерений.