Ksenija Bosnar-Vukić
Fakultet za fizičku kulturu, Zagreb

Prethodno priopćenje

ALGORITAM I PROGRAM ZA UTVRĐIVANJE KANONIKČIH RELACIJA LATENTNIH DIMENZIJA DVA SKUPA NEKVANTITATIVNIH PODATAKA

Predložen je algoritam i napisan program za utvrđivanje kanoničkih relacija ortogonaliziranih, i prethodno dovedenih u neku parsomonijsku poziciju, latentnih dimenzija dva skupa nominalnih varijabli, transformiranih u binarni oblik.

0. UVOD

U velikom broju znanstvenih disciplina često se prikupljaju nenumerički podaci, tj. istražuju se nominalne varijable, te se postavlja problem određivanja odnosa između dva sistema tih varijabli. U ovom je radu predložena procedura za utvrđivanje kanoničkih relacija dva skupa nekvantitativnih podataka u kojoj je baza za kanoničku analizu dobijena tako,da su za svaki pojedini skup nominalnih varijabli transformiranih u binarni oblik određene latentne dimenzije (algoritmom koga su predložili Momirović, Gredelj, Szirovica i Dobrić, 1980.), dovedene u jednu parsimonijsku poziciju i zatim ortogonalizirane.

1. TRANSFORMACIJA NEKVANTITATIVNIH PODATAKA U BINARNI OBLIK

rni oblik

Formirajmo skup nominalnih varijabli od kojih svaka može imati različiti broj kategorija. Opišimo entitete na tom skupu tako, da prisustvo obilježja koje definira neku kategoriju označimo »1«, a za sve ostale kategorije pojedine varijable dajmo oznaku »O«. Tako ćemo dobiti, za svaki entitet u svakoj varijabli, vektor čiji je jedan član jednak jedinici, a svi ostali članovi jednaki su nuli. Od tih vektora formirajmo matricu B koja će imati broj redova jednak broju entiteta i broj stupaca jednak ukupnom broju kategorija u svim varijablama. Za dva skupa nominalnih varijabli imat ćemo dvije takve matrice.

2. LATENTNE DIMENZIJE

Operacijom

$$C = B^T B$$

dobit ćemo matricu C čije su vandijagonalne submatrice (sačinjene od kategorija dviju varijabli) kontigencijske tabele među varijablama,dok dijagonalne submatrice (koje čine kategorije jedne varijable) u svojim dijagonalama sadrže elemente jednake broju entiteta u pojedinoj kategoriji (marginalne frekvencije), a ostali su im elementi jednaki nuli.

Dijeljenjem matrice C s N=broj entiteta

$$P = C N^{-1}$$

dobit ćemo matricu vjerojatnosti u čijoj će se dijagonali nalaziti probabiliteti pripadanja jednoj kategoriji, a u vandijagonalnim submatricama nalazit će se probabiliteti intersekcije.

Rješavanjem sistema karakterističnih jednadžbi

$$(P - \lambda_p I) X_p = 0$$

$$p = 1, \ldots, q \quad (q = \text{broj kategorija})$$

odredit ćemo svojstvene vrijednosti λ i njima pridružene svojstvene vektore X matrice P.

Kako sve vrijednosti ne moraju nositi releventne informacije, odredit ćemo kriterij za izbor svojstvenih vektora. U ovom slučaju ne postoje hipoteze o broju latentnih dimenzija, pa ćemo zadržati svojstvene vrijednosti koje su jednake ili veće od prosjeka nenultih svojstvenih vrijednosti matrice P i njima pridružene svojstvene vektore. Neka nam je λ matrica t zadržanih svojstvenih vrijednosti i neka nam je X matrica t zadržanih svojstvenih vektora.

Reducirana matrica glavnih komponenata K bit će

$$K = BX$$

Da bismo zadovoljili zahtjevu za jednostavnom strukturom, tj. da bi postigli jedostavnost opisa kategorija nominalnih varijabli u prostoru što ga razapinju vektori iz X, naći ćemo ortonormalnu matricu T koja nakon transformacije

$$XT = G = (g_{pr})$$

maksimizira varimax kriterij

$$\sum_{r=1}^{t} \left(q \sum_{p=1}^{q} g_{pr}^4 - \left(\sum_{p=1}^{q} g_{pr}^2 \right)^2 \right) q^{-2} = \max.$$

Projekcije entiteta na reducirane glavne komponente transformirane u varimax poziciju dobit ćemo operacijom

$$H = KT$$

Kao posljedica transformacije dimenzije iz H nisu više ortogonalne, osim u slučaju da su T ili λ matrice identiteta, jer su njihove relacije jednake

$$M = T^T \lambda T.$$

Ako sa Δ^2 označimo diag M, matrica strukture (tj. ortogonalne projekcije kategorija naminalnih varijabli na latentne dimenzije) bit će

$$F = X \lambda T \Delta^{-1}$$

a matrica sklopa (tj. njihove koordinate)

$$A = G \Delta.$$

3. ORTOGONALIZACIJA LATENTNIH DIMENZIJA

Standardiziranjem vrijednosti iz H (tako da je $E=(0)$, $E_2=(1)$) dobit ćemo matricu Z, korelaciju među dimenzijama bit će

$$R=Z^TZN^{-1}$$

Neka je L matrica svojstvenih vrijednosti i W matrica svojstvenih vektora matrice R. Operacijom

$$WL^{-1/2}W=R^{-1/2}$$

dobit ćemo matricu R koja ortogonalizira dimenzije iz H zadovoljavajući kriterij najmanjih kvadrata, tj.

$$ZR^{-1/2}Q$$

daje ortogonalne latentne dimenzije u Q za koje vrijedi, ako je q_{ij} rezultat entiteta e_i na komponenti Q_j formula

$$d^2=\sum_{i=1}^{N}(Z_{ij}-q_{ij})^2=\min.$$

Provođenjem istog postupka za svaki od dva skupa nominalnih varijabli dobićemo matricu Q_1 za prvi i matricu Q_2 za drugi skup.

4. KANONIČKA ANALIZA

Neka je

$$G_{12}=Q_1^TQ_2N^{-1}$$

matrica korelacije latentnih dimenzija dva skupa za koju vrijedi

$$G_{21}=G_{12}^T.$$

U matrici

$$F=G_{21}G_{12}$$

nalazit će se kvorijance dijelova latentnih varijabli drugog skupa koje sadrže varijancu predvidivu na osnovi latentnih varijabli prvog skupa.

Neka je D matrica svojstvenih vrijednosti i Y matrica svojstvenih vektora matrice F. U matrici D nalazit će se varijance dijelova latentnih varijabli drugog skupa predvidive na osnovi latentnih varijabli prvog skupa, tj. kvadrati koeficijenata kanoničke korelacije. U matrici Y nalazit će se vektori pondera koji će dati linearni kompozit latentnih dimenzija drugog skupa s najvećom količinom predvidive varijance.

Vektori pondera prvog skupa bit će

$$U=G_{12}YD^{-1/2}$$

Neka su R_{11} i R_{22} matrice interkorelacija latentnih dimenzija u dva skupa. Matrica kanoničke strukture, u kojoj su relacije između latentnih varijabli i kanoničkih varijata prvog skupa bit će

$$S_1=R_{11}U$$

tj.

$$S_2=R_{22}Y$$

za drugi skup.

5. PROGRAM CANIBAL

Ovaj program, napisan u meta jeziku SS (Zakrajšek, Štalec, Momirović, 1974), utvrđuje kanoničke relacije ortogonaliziranih latentnih dimenzija dva skupa nekvantitativnih podataka prethodno transformiranih u binarni oblik.

Korisnik mora sam napisati sequence karticu za svaki skup podataka i onoliko variable kartica koliko je analiziranih kategorija (opisano u Zakrajšek, Štalec, Momirović, 1974).

U naredbama broj 9 i 52, Ca=1/n, n=broj entiteta.

CANIBAL

```
TEXT (TEXT=CANIBAL)
OUTPUT (DEVICE=PR)
INPUT (SCORE=BB1)
INPUT (SCORE=BB2)
CONFORM (IN1=BB1, IN2=BB2, OUT1=B1, OUT2=B2)
HEADING (TEXT=ANALIZA SETA 1)
TRANSPOSE (OLD=B1, NEW=B1T)
MULT (A=B1T, B=B1, M=C1)
PRINT (MATRIX=C1, TEXT=KONTINGENCIJSKA MATRICA SETA1)
LINEAR (A=C1, CA=XXX, M=P1)
PRINT (MATRIX=P1, TEXT=MATRICA VJEROJATNOSTI SETA 1)
DIAGONALISATION (R=P1, X=X1, LAMBDA=L1)
HOTELLING (X=X1, LAMBDA=L1, F=HT1)
DELETE (MATRIX=X1)
TRANSPOSE (OLD=HT1, NEW=H1T)
MULT (A=HT1, B=H1T, M=HH1)
DIAG (A=HH1, C=-0.5, D=D1)
```

```
MULT (A=H1T, B=D1, M=X1)
PRINT (MATRIX=X1, TEXT=SVOJSTVENI VEKTORI SETA 1)
PRINT (MATRIX=H1T, TEXT=GLAVNE OSOVINE SETA 1)
MULT (A=B1, B=X1, M=K1)
PRINT (MATRIX=K1, TEXT=GLAVNE KOMPONENTE SETA 1)
TRANSPOSE (OLD=X1, NEW=X1T)
VARIMAX (F=X1T, TAU=TAU1, FN=FN1)
TRANSPOSE (OLD=FN1, NEW=FN1T)
PRINT (MATRIX=FN1T, TEXT=ROTIRANI SVOJSTVENI VEKTORI SETA 1)
TRANSPOSE (OLD=TAU1, NEW=TAU1T)
PRINT (MATRIX=TAU1T, TEXT=TRANSFORMACIJSKA MATRICA)
MULT (A=TAU1, B=HH1, M=MM1)
SCALE(C=M1, R=SM1)
PRINT (MATRIX=SM1, TEXT=KOSINUSI LATENTNIH DIMENZIJA SETA 1)
DIAG (A=M1, C=0.5, D=KD1)
DIAG (A=KD1, C=-1.0, D=DD1)
MULT (A=X1, B=HH1, M=XH1)
MULT (A=HX1, B=TAU1T, M=XHT1)
DELETE (MATRIX=XH1)
MULT (A=XT1, B=DD1, M=F1)
MULT (A=FN1T, B=KD1, M=SK1)
PRINT (MATRIX=SK1, TEXT=MATRICA SKLOPA SETA 1)
PRINT (MATRIX=F1, TEXT=MATRICA STRUKTURE SETA 1)
MULT A=K1, B=TAU1T, M=PE1)
PRINT (MATRIX=PE1, TEXT=PROJEKCIJE ENTITETA NA KOMPONENTE SETA 1)
STATISTICS (SCORE=PE1, Z=Z1)
PRINT (MATRIX=Z1, TEXT=STANDARDIZIRANE VRIJEDNOSTI IZ PE1)
CORELATION (SCORE=Z1, R=R1)
PRINT (MATRIX=R1, TEXT=INTERKORELACIJE Z1)
DIAGONALISATION (R=R1, X=XR1, LAMBDA=LR1)
ORTHOSCORES (Z=Z1, X=XR1, LAMBDA=LR1, V=Q1)
PRINT (MATRIX=Q1, TEXT=ORTOGONALIZIRANE Z1)
HEADING (TEXT=ANALIZA SETA 2)
TRANSPOSE (OLD=B2, NEW=B2T)
MUTL (A=B2T, B=B2, M=C2)
PRINT (MATRIX=C2, TEXT=KONTINGENCIJSKA MATRICA SETA 2)
LINEAR (A=C2, CA=XXX, M=P2)
PRINT (MATRIX=P2, TEXT=MATRICA VJEROJATNOSTI SETA 2)
DIAGONALISATION (R=P2, X=X2, LAMBDA=L2)
HOTELLING X=X2, LAMBDA=L2, F=HT2)
DELETE (MATRIX=X2)
TRANSPOSE (OLD=HT2, NEW=H2T)
MULT (A=HT2, M=HH2, B=H2T)
DIAG (A=HH2, C=-0.5, D=D2)
MULT (A=H2T, B=D2, M=X2)
PRINT (MATRIX=X2, TEXT=SVOJSTVENI VEKTORI SETA 2)
PRINT (MATRIX=H2T, TEXT=GLAVNE OSOVINE SETA 2
PRINT (MATRIX=K2, TEXT=GLAVNE KOMPONENTE SETA 2)
MULT A=B2, B=X2, M=K2)
PRINT (MATRIX=K2, TEXT=GLAVNE KOMPONENTE SETA 2)
TRANSPOSE (OLD=X2, NEW=X2T)
VARIMAX (F=X2T, TAU=TAU2, FN=FN2)
TRANSPOSE (OLD=FN2, NEW=FN2T)
PRINT (MATRIX=FN2T, TEXT=ROTIRANI SVOJSTVENI VEKTORI SETA 2)
TRANSPOSE (OLD=TAU2, NEW=TAU2T)
PRINT (MATRIX=TAU2T, TEXT=TRANSFORMACIJSKA MATRICA SETA 2)
MULT (A=TAU2, B=HH2, M=MM2)
MULT (A=MM2, B=TAU2T, M=M2)
SCALE (C=M2, R=SM2)
PRINT (MATRIX=SM2, TEXT=KOSINUSI LATENTNIH DIMENZIJA SETA 2)
DIAG (A=M2, C=0.5, D=KD2)
DIAG (A=KD2, C=-1.0, D=DD2)
```

```
MULT (A=X2, B=HH2, M=XH2)
MULT (A=XH2, B=TAU2T, M=XHT2)
DELETE (MATRIX=XH2)
MULT (A=XHT2, B=DD2, M=F2)
MULT (A=FN2T, B=KD2, M=SK2)
PRINT (MATRIX=SK2, TEXT=MATRICA SKLOPA SETA 2)
PRINT (MATRIX=F2, TEXT=MATRICA STRUKTURE SETA 2)
MULT (A=K2, B=TAU2T, M=PE2)
PRINT (MATRIX=PE2, TEXT=PROJEKCIJE ENTITETA NA KOMPONENTE SETA 2)
STATISTICS (SCORE=PE2, Z=Z2)
PRINT (MATRIX=Z2, TEXT=STANDARDIZIRANE VRIJEDNOSTI IZ PE2)
CORRELATION (SCORE=Z2, R=R2)
PRINT (MATRIX=R2, TEXT=INTERKORELACIJE Z2)
DIAGONALISATION (R=R2, X=XR2, LAMBDA=LR2)
ORTHOSCORES (Z=Z2, X=XR2, LAMBDA=LR2, V=Q2)
PRINT (MATRIX=Q2, TEXT=ORTOGONALIZIRA NE Z2)
HEADING (TEXT=KANONIČKA ANALIZA SETA1 I 2)
CROSSCORRELATION (P1=Q1, P2=Q2, R12=R12)
TRANSPOSE (OLD=R12, NEW=R21)
PRINT (MATRIX=R12, TEXT=KROSKORELACIJASETA 1 I 2)
MULT (A=R21, B=R12, M=FF)
DIAGONALISATION (R=FF, X=Y2, LAMBDA=LF)
TRANSPOSE (OLD=Y2, NEW=Y2T)
MULT (A=Y2, B=FF, M=Y2F)
MULT (A=Y2F, B=Y2T, M=YFYT)
DELETE (MATRIX=Y2F)
DIAG (A=YFYT, D=CRR)
DIAG (A=CRR, C=0.5, D=CR)
DIAG (A=CR, C=-1.0, D=CRI)
PRINT (MATRIX=CR, TEXT=KANONIČKE KORELACIJE)
MULT (A=R12, B=Y2T, M=RY)
A=RY, B=CRI, M=Y1)
PRINT (MATRIX=Y1, TEXT=PONDERE ZA SET 1)
PRINT (MATRIX=Y2T, TEXT=PONDERE ZA SET 2)
DELETE (MATRIX=RY)
MULT (A=R1, B=Y1, M=CF1)
MULT (A=R2, B=Y2T, M=CR2)
PRINT (MATRIX=CF1, TEXT=KANONIČKA STRUKTURA SETA 1)
PRINT (MATRIX=CF2, TEXT=KANONIČKA STRUKTURA SETA 2)
```

LITERATURA

1. Momirović, K., M. Gredelj, V. Dobrić i L. Szirovica: BURT-algoritam i program za određivanje latentnih dimenzija skupa nekvantitativnih podataka. Rukopis, SRCE, 1980.
2. Momirović, K., i L. Szirovica, M. Gredelj i V. Dobrić: Latentna struktura nominalnih varijabli. Rukopis, SRCE, 1980.
3. Zakrajšek, E., J. Štalec i K. Momirović: SS-programski sistem za multivarijatnu analizu podataka. Zbornik simpozija »Kompjuter na Sveučilištu«, C8 - 1 - c8 - 16, 1974., također i u Zborniku simpozija»Primjena kompjutera u psihološkim istraživanjima«, Skopje, 1975.
4. Fulgozi, A.: Faktorska analiza, Školska knjiga, Zagreb, 1979.

THE ALGORHYTHM AND THE PROGRAMME DESIGNED TO EFTABLISH CANONICAL RELATIONS BETWEEN LATENT DIMENSIONS OF TWO GROUPS OF NON-QUANTITATIVE DATA

The algorythm is proposed and the programe written to establish canonical relations between latent dimensions of two groups of nonquantitative data transformed into binary form. The basis for canonical analysis is obtained by means of orthogonalization of entity values on the reduced major components of the probability matrix, transformed into varimax position. The programe is written in the SS meta-language (Zakrajšek, Štalec, Momirović, 1974).

АЛГОРИТМ И ПРОГРАММА ДЛЯ ОПРЕДЕЛЕНИЯ КАНОНИЧЕСКИХ СООТНОШЕНИЙ МЕЖДУ ЛАТЕНТНЫМИ ДВУХ ГРУПП НЕКВАНТИТАТИВНЫХ ДАННЫХ

Предлагаются алгоритм и программа для определения канонических соотношений между латентными измерениями двух групп неквантитативных данных, трансформированных в бинарную форму. Основа для канонического анализа получена при помощи ортогонализации величин заданий на редуцированных главных компонентах матрицы вероятностей, трансформированных в варимакс позицию. Программа составлена в мета--языке СС (Закрайшек, Шталец, Момирович, 1974 г.)