188

Konstantin Momirović

Katedra za kineziološku
psihologiju i sociologiju

Egon Zakrajšek

Inštitut za matematiko, fiziko in
mehaniko Univerze v Ljubljani

ODREĐIVANJE TAKSONOMSKIH SKUPINA DIREKTNOM OBLIMIN TRANSFORMACIJOM ORTOGONALIZIRANIH ORIGINALNIH I LATENTNIH VARIJABLI

DETERMINING TAXONOMIC GROUPS BY DIRECT OBLIMIN TRANSFORMATION OF ORTHOGONALISED ORIGINAL AND LATENT VARIABLES

The real possibility for generating adequate transforming operators exists only if relatively homogeneuous groups are determined and operators are generated according to the initial state vector of every group and the ideal final state vector of that group.

Because of some disadvantages of taxonomic methods which were proposed by Ward, Sokal, Cattell, Johnson, Friedman i Rubin, Hartigan, Momirović, Wee, Fukunaga and Koontz, Zahn and others, the ne methods is developed which determines taxonomic groups using direct oblimin transformation of original vectors, orthogonalised according to the least squares criterion, transformed to principal components od orthogonalised in any other way. Orthogonalised latent dimensions can be used as the basis of taxonomic method.

The essence of the method if transformation of normally distributed variables to the variables having Weibull distribution, with Weibull's parameter m as low as possible.

The method is suitable to determine objectively and invariantly taxonomic groups as well as relations between these groups and relations between each entity and each of isolated group. The method also provides us with possibilities to determine directly the taxonomic group characteristic in manifest or latent space.

ОПРЕДЕЛЕНИЕ ТАКСОНОМИЧЕСКИХ ГРУШ ПРИ ПОМОЩИ ПРЯМОЙ ОБЛИМИН ТРАНСФОРМАЦИИ ОРТОГОНАЛИЗИРОВАННЫХ ОРИГИНАЛЬНЫХ И ЛАТЕНТНЫХ ИЗМЕНЯЕМЫХ

Реальная возможность порождения адекватных трансформационных операторов требует определения относительно гомогеной таксономической группы и порождения операторов в соответствии с вектором начального состояния каждой из групп и с желанным вектором конечного состояния этой группы. По причине определенных недостатков таксономичеких методов, которые предлагали Ward, Sokal, Cattell, Johnson, Friedman u Rubin, Hartigan, Momirović, Wee, Fukunaga и Koontz, Zahn и др., составлен метод, определяющий таксономические группы при помощи прямой облимин трансформации оригинальных векторов, которые ортогонализированы в соответствии с критерием самых маленьких квадратов, трансформированных в главные компоненты или ортогонализированных каким-либо другим способом.
Суть метода сосотоит в трансформации нормально распределенных оригинальных изменяемых в изменяемые, обладающие распределением Weibull-a с возможно более низким параметром.

Метод позволяет объективное и инвариантное определение таксономических групп, определение соотношений между этими группами и соотношений между каждой единицей и каждой изолированной группой. Метод также дает возможность определения характеристик таксономических групп в манифестном и латентном пространстве.

1. UVOD

Realna mogućnost za generiranje adekvatnih transformacijskih operatora postoji samo ako se utvrde relativno homogene taksonomske skupine i operatori generiraju u skladu sa vektorom inicijalnog stanja svake skupine i željenim vektorom finalnog stanja te skupine.

Iako je, osobito u posljednjih nekoliko godina, predloženo više taksonomskih procedura (Cattell, 1966; Johnson, 1967; Friedman i Rubin, 1967; Hartigan, 1967; Momirović, 1968, 1969, 1972; Ward, 1963; Wee, 1968, 1971; Wishart, 1968; Fukunaga i Koontz, 1970, 1970; Zahn, 1971; Koontz i Fukunaga, 1972) ni jedna od njih nije bez nedostataka promatrana pod vidom primjene u kineziologijskim istraživanjima. Dva su problema koja nisu dovoljno uspješno riješena u dosadašnjim taksonomskim procedurama: određivanje broja taksona i invarijantnost rezultata dobijenih na osnovu različitih početnih operatora.

Metoda koja će biti predložena u ovom radu nema ni jedan od tih nedostataka, ali to nipošto ne znači, da je stvarno efikasna kao taksonomska procedura. Ona omogućava, prije svega, veću širinu izbora onima, koji, zbog znanstvenih ili praktičnih razloga, moraju taksonomizirati neki skup entiteta, i vjerojatno da će, u nekim slučajevima biti pogodnija od drugih metoda taksonomske analize.

2. PRINCIPI METODE

Taksoni se očito ne mogu identificirati u prostoru što ga omeđuje sistem multivarijatno normalno distribuiranih varijabli. Ni ako su samo marginalne distribucije tog sistema normalne, taksoni se ne mogu, u pravilu, izolirati, a nikako ne ako su relacije marginalnih varijabli linearne.

Jedna od mogućnosti za određivanje taksona je takva linearna transformacija čitavog skupa varijabli, koja svaku originalnu varijablu pretvara u varijablu koja ima Weibullovu distribuciju.

$$F(X) = KX^{m} \exp KX^{(m+1)} / (m+1)$$

$$m = 0,1,\dots$$

gdje je K parametar za koga vrijedi

$$E(X) = (K/(m+1))^{1/(m+1)} \ \gamma \ ((m+2)/(m+1))$$
$$(K/(m+1))^{2/(m+1)}$$

a γ je gamma funkcija. Ova transformacija, pri tome, treba da zadovolji i uvjet, za sistem varijabli X_j, X_k ; $j,k = 1,n$ i entiteta S_i , $i = 1,N$.

$$\text{IF } P(X_{ij}) \angle \varepsilon \ , \ P(X_{ik}) \geqslant \varepsilon \ ; \ k=1,n \ ; \ j \neq k$$

gdje je P oznaka vjerojatnosti rezultata X_{ij}, X_{ik} a $O < \varepsilon < 1$.

Očito, što je parametar m Weibullove distribucije niži, ovaj je uvjet lakše postići.

Relativno jednostavan način da se ovo postigne je primjena poznatog Carrollovog postupka za objektivnu kosu transformaciju nekog koordinatnog sistema u parsimonijsku poziciju. Ovdje je primijenjena Jennrich-Sampsonova direktna oblimin procedura, u modifikaciji E. Zakrajšeka, koja znatno ubrzava inače spori iterativni proces, na originalnu matricu podataka, standardiziranih i zatim ortogonaliziranih na različite načine, kao i na matricu ortogonaliziranih latentnih dimenzija koje su dovedene, prethodno, u neku parsimonijsku poziciju.

3. ORTOGONALIZACIJA KOMPLETNOG SKUPA ORIGINALNIH VARIJABLI

Neka je

$$Z = (z_{ij}) \qquad i = 1,\dots,N \quad j = 1,\dots,n \quad (N \quad n)$$

matrica standardiziranih i normaliziranih rezultata N entiteta S_i u n varijabli V_j . Neka je

$$R = Z'Z \ N^{-1}$$

matrica interkorelacija varijabli V_j . Ako su varijable linearno nezavisne, R će biti pozitivno definitna matrica. Neka su

$$X = (x_{jp}) \qquad j = 1,\dots,n \quad p = 1,\dots,n$$

i

$$L = (L_p) \qquad p = 1,\dots,n$$

matrice karakterističnih vektora i karakterističnih korjenova od R, dobijene rješavanjem karakterističnih jednažbi

$$(R - L_p \ I) \ x_p = O \qquad p = 1,\dots,n$$

Operacijom

$$XL^{-1/2} X' = R^{-1/2}$$

dobit će se matrica $R^{-1/2}$, koja, zadovoljavajući kriterij najmanjih kvadrata, ortogonalizira originalne varijable V_j . Naime, operacija

$$ZR^{-1/2} = W$$

proizvodi ortogonalne varijable W_j , za koje vrijedi, ako je w_{ij} rezultat entiteta S u varijabli W_j

$$d_j^2 = \sum_{i=1}^{N} (z_{ij} - w_{ij})^2 = \min \qquad j = 1,\dots,n$$

Ovo istraživanje provedeno je u okviru programa »Utjecaj tjelesne aktivnosti na psihosomatski status« u dijelu koji je usmjeren na rješavanje nekih problema koji omogućavaju efikasno generiranje kineziologijskih transformacijskih operatora. Istraživanje je financirao Savjet za naučni rad SR Hrvatske na temelju ugovora IX/16 od 19. 6. 1972. godine.

Matrica W je prema tome pogodna ortogonalna osnova za TAXOBL proceduru, ako je zbog nekog razloga potrebno da se zadrže sve početne informacije pri određivanju taksonomskih grupa.

Naravno, u istu svrhu može poslužiti i bilo koja ortogonalna transformacija matrice W. Neka je T bilo koja ortonormalna matrica reda n. Svaka matrica, dobijena operacijom

$$WT = V = (v_{ij})$$

uz uslov da uvijek vrijedi $TT' = T'T = I$, može biti osnov za TAXOBL proceduru. Posebno je pogodna transformacija kod koje matrica V zadovoljava varimax kriterij

$$N = \sum_{j=1}^{n} \sum_{i=1}^{N} v_{ij}^4 - \sum_{j=1}^{n} \left(\sum_{i=1}^{N} v_{ij}^2\right)^2 = \max$$

jer to znatno ubrzava konvergenciju TAXOBL procedure.

Još nekoliko metoda za ortogonalizaciju originalnog skupa varijabli može dati pogodne ortogonalne osnove za ovu taksonomsku proceduru. Najprikladnija je, vjerojatno, transformacija čitavog skupa originalnih varijabli u glavne komponente, tj.

$$ZXL^{1/2} = \Omega$$

odnosno

$$ZXL^{-1/2} = \omega$$

ako je podesnije da nove varijable imaju jednake varijance.

I jednostavna ortogonalizacija square-root metodom može biti primijenjena, ako je sporedno kakva je osnova za određivanje taksona, već je dovoljno sačuvati sve informacije sadržane u originalnim varijablama. U tom slučaju, ortogonalna osnova za TAXOBL bit će

$$\Phi = Zt'^{-1}$$

ako je t matrica, dobijena metodom Choleskog, za koju vrijedi

$$R = tt'$$

jer, očito

$$\Phi'\Phi N^{-1} = t^{-1} Rt'^{-1} = I$$

I svaka ortogonalna transformacija matrice t može poslužiti u istu svrhu.

4. ORTOGONALNE LATENTNE DIMENZIJE KAO OSNOV ZA TAXOBL PROCEDURE

Ako je potrebno da se zadrže samo valjane informacije pri određivanju taksonomskih grupa, vjerojatno je najrazboritije zadržati samo onoliko linearnih kombinacija originalnih varijabli, koliko je upravo potrebno da se obuhvati donja granica veličine valjane varjance sistema varijabli V_j, tj.

$$v = tr\,(I - dg^{-1} R^{-1})$$

tako da je broj tih dimenzija, k, definiran uvjetima

$$\sum_{p=1}^{k} L_p \geqslant v$$

$$\sum_{p=1}^{k-1} L_p \angle v$$

Ako je potrebno da se dostigne granica maksimalne valjane varijance, k je vjerojatno najpouzdanije definiran kao

$$k = NUM\,(L_p \geqslant 1.)$$

Kako god definiran bio k, neka je

$$\breve{X} = (x_{jp})$$

$$j = 1, \ldots, n$$
$$p = 1, \ldots, k$$

i

$$\breve{L} = (L_p)$$

Operacija

$$Z\breve{X}\breve{L}^{-1/2} = K$$

dat će ortogonalni sistem varijabli K_p, čije su varijance reducirane na 1. Sistem K može također biti pogodna ortogonalna osnova za TAXOBL proceduru. Međutim, u nekim će slučajevima biti podesnije da varijanca linearnih kombinacija varijabli Z_j bude nestandardizirana, tj. ravna L_p; u tom slučaju, naravno, operacija

$$Z\breve{X}\breve{L}^{1/2} = J$$

proizvodi varijable J_p sa varijancama L_p. Očito je, međutim, da su varijable K_p i J_p kolinearne.

U većini slučajeva bit će međutim pogodnije, da osnov za TAXOBL proceduru bude neka aproksimacija realnih latentnih dimenzija.

Neka je

$$H = \breve{X}\breve{L}^{1/2}$$

i neka je Q neka matrica koja, uz uslov

$$dg(QQ') = I$$

omogućava transformaciju

$$HQ^{-1} = A = (a_{jp})$$

pri čemu elementi a_{jp} matrice A zadovoljavaju uslov

$$\sum_{p}^{k} \sum_{< q}^{k} \sum_{j=1}^{n} a_{jp}^2\, a_{jq}^2 = \min$$

Neka je

$$QQ' = M$$

i neka je

$$Y'MY = D$$

gdnje su Y i D matrice karakterističnih vektora, odnosno korjenova od M. Sada je

$$YD^{½}\ Y' = M^{½}$$

pa operacija

$$G = AM^{½}$$

daje ortogonalnu aproksimaciju, pod vidom kriterija najmanjih kvadrata, latentnih dimenzija definiranih kao oblimin faktori. Ako je

$$U^2 = I - dg(GG')$$

Anderson-Rubinova operacija

$$ZU^{-2}\ G(G'U^{-2}\ RU^{-2}\ G)^{-½} = C$$

proizvest će ortogonalne varijable C_p, koje uz uslov

$$C'CN^{-1} = I$$

zadovoljavaju i uslove

$$d_p^2 = \sum_{i=1}^{N} (c_{ip} - p_{ip})^2 = \min$$

$$p = 1, \ldots, k$$

ako su p_{ip} elementi matrice oblimin vrijednosti entiteta S_i, dobijene operacijom

$$P = ZR^{-1}F$$

gdje je

$$F = AM$$

matrica korelacija originalnih varijabli i oblimin faktora.

5. ALGORITAM ZA TAXOBL PROCEDURE TAKSONOMSKE ANALIZE

Neka je B bilo kakva ortogonalna osnova za TAXOBL proceduru, tj. neka je B bilo koja od matrica Ω, ω, Φ, W, V, K, J, G. Carrollov oblimin kriterij u suštini je Weibullizacija distribucije rezultata u vektorima b_j matrice B. Otuda, transformacijska matrica S, koja, uz uslov

$$dg(SS') = I$$

operacijom

$$BS^{-1} = E = (e_{ij})$$

proizvodi matricu E čiji elementi zadovoljavaju oblimin kriterij

$$\sum_{j<k}^{n} \sum^{n} \sum_{i=1}^{N} e_{ij}^2\ e_{ik}^2 = \min$$

$$j,k = 1, \ldots, n$$
$$j < k$$
$$i = 1, \ldots, N$$

proizvodi i taksonomsku raspodjelu entiteta S_i.

U stvari, TAXOBL algoritam ima dvije faze.

U prvoj, operacijom

$$BS_o = E_o$$

gdje matrica S_o zadovoljstva, uz uslov ortonormalnosti

$$S'_o\ S_o = S_o\ S'_o = I$$

ortogonalni oblimin kriterij

$$\sum_{j<k}^{n} \sum^{n} \sum_{i=1}^{N} e_{ij\,(o)}^2\ e_{ik\,(o)}^2 = \min$$

provizorno se zarotira matrica B u ortogonalnu matricu sličnu finalnoj matrici E, pa se zatim, operacijom

$$E_o\ S_f^{-1} = E$$

dobije konačna solucija. Naravno, transformacijska matrica S je

$$S = S_o\ S_f^{-1}$$

Matrica S je, u stvari, matrica kosinusa između ortogonalne baze B i vektora taksona. Relacije između originalnih varijabli i taksona, u ovisnosti od toga kako je ortogonalizirana ili kondenzirana matrica Z, mogu se dobiti operacijama

$g_1 = XL^{½}S$, ako je $B = ZXL^{½}$

$g_2 = XL^{-½}S$, ako je $B = ZXL^{-½}$

$g_3 = t'^{-1}S$, ako je $B = Zt'^{-1}$

$g_4 = R^{-½}S$, ako je $B = ZR^{-½}$

$g_5 = R^{-½}TS$, ako je $B = ZR^{-½}T$

$g_6 = \tilde{X}\tilde{L}^{-½}S$, ako je $B = Z\tilde{X}\tilde{L}^{-½}$

$g_7 = \tilde{X}\tilde{L}^{½}S$, ako je $B = Z\tilde{X}\tilde{L}^{½} = ZH$

$g_8 = U^{-2}\,G\,(G'\,U^{-2}\,RU^{-2}\,G)^{-½}\,S$, ako je
$B = ZU^{-2}\,G(G'\,U^{-2}\,RU^{-2}\,G)^{-½}$

Svakako, da matrica S ima posebnog smisla ako je kao osnov za taksonomsku analizu upotrebljen sistem latentnih dimenzija. Relacija između taksona i realnih latentnih dimenzija su

$$g = F'U^{-2}\,FM^{-½}\,(M^{-½}F'U^{-2}\,RU^{-2}\,FM^{-½})^{½}S$$
$$= F'U^{-2}\,G(G'U^{-2}\,RU^{-2}\,G)^{-½}S$$

Relacije između taksona, definirane kosinusima kutova što ih zaklapaju vektori taksona, mogu se dobiti operacijom

$$SS' = \eta$$

Ortogonalne projekcije entiteta na taksonomske vektore su

$$X = E\eta$$

i mogu biti tretirane kao taksonomske varijable.

Vjerojatno je najpogodnije, da se za identifikaciju entiteta S_i kao člana taksona T_p primijeni jednostavni algoritam

$$\text{IF } e_{ip} \max(p),\ e_{ip} = 1$$
$$\text{IF } e_{ip} \neq 1,\ e_{ip} = 0$$
$$p = 1, \ldots, t$$

pa binarna matrica Ψ, sačinjena od ovako transformiranih koeficijenata e_{ip} označava pripadnost svakog entiteta S_i, $i=1,\ldots, N$ taksonima T_p, $p = 1, \ldots, t$.

6. ZAKLJUČAK

Realna mogućnost za generiranje adekvatnih transformacijskih operatora postoji samo ako se odrede relativno homogene taksonomske skupine i operatori generiraju u skladu sa vektorom inicijalnog stanja svake skupine i željenim vektorom finalnog stanja te skupine.

Zbog izvjesnih nedostataka taksonomskih metoda koje su predložili Ward, Sokal, Cattell, Johnson, Friedman i Rubin, Hartigan, Momirović, Wee, Fukunaga i Koontz, Zahn i drugi, izrađena je metoda koja taksonomske skupine određuje direktno oblimin transformacijom originalnih vektora, ortogonaliziranih u skladu s kriterijem najmanjih kvadrata, transformiranih u glavne komponente ili ortogonaliziranih na bilo koji drugi način. Kao baza taksonomskog postupka mogu biti upotrebljene i ortogonalizirane latentne dimenzije.

Bit metode je transformacija normalno distribuiranih originalnih varijabli u varijable koje imaju Weibullovu distribuciju, koja ima što je moguće niži Weibullov parametar m.

Metoda omogućava objektivno i invarijantno određivanje taksonomskih skupina, relacija između tih skupina i relacija svakog entiteta sa svakom od izoliranih skupina. Metoda također omogućava direktno određivanje karakteristika taksonomskih skupina u manifestnom ili latentnom prostoru.

7. LITERATURA

1. Cattell, R. B.
Handbook of multivariate experimental psychology. Rand, Chicago, 1966.

2. Dorn, W. S. i H. J. Greenberg
Mathematics and computing: with FORTRAN programming. Willey, New York, 1967.

3. Fukunaga, K. i W. L. G. Koontz
Application of the Karhunen-Loeve expansion to feature selection and odreding. IEEE Transactions on Computers, 1970. C-19, 4, pp. 311-318.

4. Fukunaga, K. i W. L. G. Koontz
A criterion and an algorithm for grouping data. IEEE Transactions on Computers, 1970. C-19, 10, pp. 917-923.

5. Friedman, H P. i J. C. Rubin
On some invariant criteria for grouping data. Journal of the American Statistical Association, 1967. 62, pp. 1159-1178.

6. Harman, H. H.
Modern factor analysis. 2. ed. The University of Chicago Press, Chicago, 1970.

7. Hartigan, J. A.
Representation of similarity matrices by tries. Journal of the American Statistical Association, 1967. 62, pp. 1140-1158.

8. Jennrich, R. I. i P. F. Samson
Rotation for simple loadings. Psychometrika, 1966. 31, 3, pp. 313-323.

9. Johnson, S. C.
Hierarchacal clustering shemes. Psychometrika, 1967. 32, 3, pp. 241-254

10. Koontz, W. L. G. i K. Fukunaga
A nonparametric valley-seeking technique for cluster analysis. IEEE Transactions on Computers, 1972. C-21, pp. 171-178.

11. Lee, J. A. N.
Numerical Analysis for Computers. Reinhold, New York, 1966.

12. Momirović, K.
Neke ralacije između diskriminativne kanoničke analize i modificirane Q-metode faktorske analize. Psihologija, 1968. 1, 2, str. 173-175.

13. Momirović, K.
Prilagane na obektivni tipologi''ni proceduri pri orientiraneto i podbora na sportistite i pri programiraneto na trenirovkite. U »Aktualni problemi na psihologijata na sporta« Sofija, 1968. str. 199-209.

14. Momirović, K.
Određivanje psiholoških tipova iterativnom primjenom modificirane Q-metode faktorske analize. Psihologija, 1969. 2, 1, str. 343-346.

15. Momirović, K.
Metode za transformaciju i kondenzaciju kinezioloških informacija. Institut za kineziologiju, Zagreb, 1972.

16. Orchard — Hays, W.
Advanced Linear-programming Computing Techniques. McGraw-Hill, New York, 1968.

17. Sokal, R. R. i P. H. A. Sneath
Principles of Numerical Taxonomy. Freeman, San Francisko, 1963.

18. Štalec, J. i K. Momirović
Ukupna količina valjane varijance kao osnov kriterija za određivanje broja značajnih glavnih komponenata. Kineziologija, 1971. 1, 1, str. 79-81.

19. Ward, J. H.
Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 1963. 58, pp. 236-244.

20. Wee, W. G.
Generalized inverse approach to adaptive multiclass pattern classification. IEEE Transactions on Computers, 1968. C-17, 12, pp. 1157-1164.

21. Wee, W. G.
Generalized inverse approach to clustering, feature selection, and classification. IEEE Transactions on Information Theory, IT-17, 1971. 3, pp. 262-269.

22. Wishart, D.
Numerical classification method for deriving natural classes. Nature, 1969. 221, pp. 97-98.

23. Zahn, C. T.
Graph-theoretical methods for detecting and describing gestalt clusters. IEEE Transactions on Computers, C-20, 1, 1971. pp. 68-86.