

JANEZ ŠTALEC, KSENIJA BOSNAR, FRANJO PROT
I KONSTANTIN MOMIROVIĆ

Fakultet za fizičku kulturu Sveučilišta u Zagrebu

Prilježeno 29. 4. 1982.

ELEMENTARNA PROCEDURA ZA IDENTIFIKACIJU KINEZIOLOŠKIH TAKSONA

Prot
Bosnar

SAŽETAK

Predložen je algoritam i napisan program koji identificira hipotetske taksona u jednom skupu entiteta koji je opisan nad jednim skupom kvantitativnih varijabli. Algoritam identificira taksona na temelju mjera sličnosti entiteta sukcesivnom primjenom jedne modifikacije Q metode taksonomske analize i jedne modifikacije generalne Guttmanove procedure za faktorizaciju kvadratnih simetričnih matrica.

Osnovnom algoritmu pridružena su dva modula. Prvi određuje sklop i strukturu taksonomskih dimenzija u prostoru standardiziranih varijabli, a drugi proizvodi, pod generalnim linearnim modelom, operator prepoznavanja koji može biti primjenjen za taksonomsku alokaciju ma kog novog objekta opisanog nad identičnim skupom varijabli.

1. UVOD

U mnogim je kineziološkim skupinama moguće nekom apriornom procedurom identificirati grupe srodnih entiteta, ili barem pretpostaviti postojanje konačnog broja takvih grupa i svaki entitet pridružiti jednoj, i samo jednoj od tih grupa. Tako je, na primjer, moguće

- skup sportskih disciplina razdijeliti na podskupove na temelju prevalencije osnovnih struktura kretanja;
- skup elemenata u sportskoj gimnastici razdijeliti na podskupove biomehanički srodnih elemenata;
- skup igrača u nekoj nogometnoj momčadi razdijeliti u podskupove na temelju njihovih uloga u obrani, organizaciji igre i napadu;
- skup učenika neke škole razdijeliti u podskupove na temelju procjene njihovih sklonosti, interesa i sposobnosti da se intenzivnije bave nekim od sportova obuhvaćenih izbornim programom nastave;
- skup pacijenata za koji se programira neki kineziterapijski tretman razdijeliti na podskupove na temelju vrsta i intenziteta deformacija, oštećenja ili oboljenja;
- skup radnika neke tvornice u kojoj se planira uvođenje sustavnih rekreativnih aktivnosti razdijeliti na podskupove na temelju procjene mogućih negativnih efekata poslova i zadataka koje ti radnici obavljaju na svojim radnim mjestima.

U većini ovih slučajeva moguće je i neophodno skup entiteta opisati i nad nekim skupom kvantitativnih varijabli, značajnih za programiranje i kontrolu efekata kineziološkog tretmana ili za objektivnu deskripciju stanja i/ili promjena sustava koji je predmet analize. U takvim je slučajevima od presudne važnosti utvrditi da li apriorna taksonomija entiteta odgovara taksonomiji u prostoru varijabli kojim su entiteti opisani, kakav je sklop i kakve su relacije taksonomskih dimenzija u prostoru varijabli, te da li je, uz razumno malu pogrešku, moguće odrediti taksonomsku alokaciju ma kog novog entiteta opisanog nad istim skupom varijabli.

U tu svrhu predložen je algoritam ONIKEN i napisan program istog imena koji identificira hipotetske taksona u jednom skupu entiteta koji je opisan nad jednim skupom kvantitativnih varijabli. Algoritam se sastoji od tri osnovna modula. Prvi modul identificira hipotetske taksona na temelju mjera sličnosti entiteta sukcesivnom primjenom jedne modifikacije Q metode taksonomske analize (Momirović, 1969; Rudan, Szivovica i Momirović, 1978, 1979; Momirović, 1980) i jedne modifikacije generalne Guttmanove procedure za faktorizaciju kvadratnih simetričnih matrica (Herak, Gredelj i Momirović, 1982; Štalec i Momirović, 1982). Drugi modul određuje sklop, strukturu i korelaciju taksonomskih dimenzija u prostoru standardiziranih varijabli postupkom analognom onome koji se, za identifikaciju latentnih dimenzija, primjenjuje u procedurama koje pripadaju području metričkog multidimenzionalnog skaliranja (Gredelj, Momirović, Dobrić i Lužar, 1982). Treći modul proizvodi, na temelju jednog algoritma za prepoznavanje uzoraka koji se osniva na generalnom linearnom modelu (Prot, Bosnar, Štalec i Momirović, 1982) operator prepoznavanja koji može biti primjenjen za taksonomsku alokaciju ma kod novog entiteta opisanog nad identičnim skupom varijabli, pod uvjetom da je skup entiteta bio dovoljno reprezentativan za stabilnu procjenu parametara modela.

2. ALGORITAM

2.1. Početne operacije

Neka je $B=(b_{ij})$, $i=1, \dots, n$; $j=1, \dots, m$ matrica podataka entiteta iz skupa $E=\{e_i; i=1, \dots, n\}$ opisanih nad kvantitativnim varijablama iz skupa $V=\{v_j; j=1, \dots, m\}$. Neka je $S=(s_{ik})$, $i=1, \dots, n$; $k=1, \dots, l$ alokacijska matrica koja apriorno određuje pripadnost entiteta $e_i \in E$ jednom od hipotetskih taksona $w_k \in \Omega = \{w_k; K=1, \dots, l\}$ tako da je $s_{ik}=1$ ako se za entitet e_i smatra da pripada taksonu w_k i $s_{ik}=0$ ako e_i ne pripada jednom hipotetskom taksonu.

Nakon reparametrizacije ulazne matrice podataka operacijom $P=BD^{-1}$, gdje je $D=\text{diag } B^T B$, mjere sličnosti entiteta moguće je definirati kao normirane skalarnе podukte vektora entiteta, tj. $C=\Delta^{-1/2} P P^T \Delta^{-1/2}$ gdje je $\Delta=\text{diag. } P P^T$.

2.2. Inicijalna solucija

Ako se hipotetska taksonomska dimenzija w_k definira sumacijom normiranih rezultata v_i svih e_i pripadnika w_k , može se pokazati da je $T_0=S^T C S$ matrica kovarijanci tako definiranih dimenzija. U dijagonalnoj matrici $\tau_0=\text{diag. } T_0$ nalaze se varijance taksonomskih dimenzija, pa će korelacije među njima biti $M_0=\tau_0^{-1/2} T_0 \tau_0^{-1/2}$. Faktorska matrica $F_0=C S \tau_0^{-1/2}$ sadrži ortogonalne projekcije entiteta na taksonomske dimenzije definirane sumacionim postupkom. Koordinate entiteta na taksonomskim dimenzijama su u matrici $A_0=F_0 M_0^{-1}$ pa su vrijednosti a_{0ij} pokazatelji pripadnosti entiteta ovako definiranim taksonomskim dimenzijama.

2.3. Finalna solucija

Ako A_0 tretiramo kao novu alokacijsku matricu, pod Q modelom taksonomske analize $P^T \Delta^{-1/2} = \Phi A_0^T + E$, gdje je Φ matrica vrijednosti varijabli na taksonima, a E matrica reziduala rezultata, matrica $\Gamma=A_0(A_0^T A_0)^{-1}$ je regresijska matrica za procjenu vrijednosti entiteta na taksonima skupa Ω . Kako je $\phi=P^T \Delta^{-1/2} \Gamma$, kovarijance taksonomskih dimenzija su u matrici

$$T = \phi^T \phi = \Gamma^T C \Gamma, \text{ s varijancama u dijagonalnoj matrici}$$

$\tau=\text{diag. } T$. Kosinusi kuteva među dimenzijama biti će $M=\tau^{-1/2} T \tau^{-1/2}$. Projekcije entiteta na taksonomske dimenzije su u matrici $F=C \Gamma \tau^{-1/2}$. Koordinate entiteta, tj. definitivni pokazatelji pripadnosti entiteta hipotetskim taksonima, nalaze se u matrici $A=FM^{-1}$.

2.4. Sklop i struktura taksonomskih dimenzija u prostoru standardiziranih varijabli

Neka je $Z=(z_{ij})$ matrica standardiziranih rezultata iz B . Struktura taksonomskih dimenzija u prostoru stan-

dardiziranih varijabli biti će $H=Z^T A$. Kako su kore-

lacije među dimenzijama jednake $R=\alpha^{-1/2}(A^T A)\alpha^{-1/2}$ $\alpha=\text{diag. } (A^T A)$, sklop taksonomskih dimenzija u prostoru standardiziranih varijabli definiran je s $G=HR^{-1}$.

2.5. Prepoznavanje taksona

Neka je matrica Y formirana tako da je matrici B pridružen jedinični vektor s n jedinica, tj. $Y=[1:B]$. Regresijski koeficijent $\beta=(Y^T Y)^{-1} Y^T A$ su operatori prepoznavanja taksona pa s operacijom $Y_p \beta$ gdje je Y_p matrica podataka entiteta iz bilo kojeg skupa E_p izvučenog iz populacije P , opisanih nad kvantitativnim varijablama skupa V) može utvrditi pripadnost entiteta po jedinom hipotetskom taksonu.

Pogreška prepoznavanja taksona može se definirati kao razlika između inicijalne i definitivne alokacije entiteta na taksonima, pa pokazatelje pogreške daje operacija $S-Y\beta$.

3. PROGRAM

Program ONIKEN napisan je u verziji 5.2/M meta jezika SS i dostupan je korisnicima Sveučilišnog računskog centra u Zagrebu. Program pretpostavlja da matrica ulaznih podataka nije reda većeg od 250. U uputama za korištenje programa pohranjenih u javnoj biblioteci SRCE SS-MAKRO nalaze se i upute o tome kako treba aktivirati makro program ONIKEN.

4. NUMERIČKI PRIMJER

Efikasnost algoritma i programa ONIKEN provjerena je na rezultatima testova dobijenim u sklopu provjere motoričkih sposobnosti 65 slučajno odabranih kandidata i kandidatkinja za studij na Fakultetu za fizičku kulturu. Pretpostavljeno je da će ispitanici, opisani nad skupom motoričkih varijabli tako da je rezultat u testovima definiran kao najbolji rezultat u više mjerenja¹, tvoriti dva taksona obzirom na spol.

Trideset i tri muškarca i trideset i dvije žene izmjereni su slijedećim testovima:

MBKPOP (provlačenje i preskakivanje)
MAGKUS (koraci u stranu)
MBAP 20 (stajanje na klupici za ravnotežu poprečno na dvije noge sa otvorenim očima)
MFLPRR (pretklon raskoračno)
MBFTAP (taping rukom)
MFESDM (skok udalj s mjesta)
MRABPT (bench press test)

Na temelju hipotetske alokacijske matrice, formirane prema spolu ispitanika, uspješno su identificirana dva taksona (označeni sa »muško« i »žensko«) definirana kao polovi jedne bipolarne dimenzije, što je vidljivo iz matrice strukture taksonomskih dimenzija u prostoru standardiziranih varijabli (tabela 1) i iz korelacije taksonomskih dimenzija koja iznosi $r=-0.998$.

Finalna alokacija ispitanika, učinjena pomoću regresijskih koeficijenata (tabela 2) pokazala je da je algoritmom i programom, s obzirom na hipotetsku alokaciju, ispravno prepoznato svih 32 ispitanice i 30 od 33 ispitanika. Inspekcijom rezultata trojice »krivo« prepoznatih ispitanika vidljivo je da su, lošim rezultatima u testovima MBKPOP, MFESDM i MRABPT, mnogo bliži parametrima ženskog uzorka. Navedeni primjer pokazuje da je algoritam i njemu pridružen program ONIKEN efikasno sredstvo za konfirmativnu analizu skupa entiteta opisanih nad skupom kvantitativnih varijabli kao i za taksonomsku alokaciju entiteta.

¹ s izuzetkom testa MRABPT, kod kojeg se mjerenje vršilo samo jednom

Tabela 1

STRUKTURA TAKSONOMSKIH DIMENZIJA

	Muško	Žensko
MBKPOP	— .48	.49
MAGKUS	— .60	.60
MBAP20	— .07	.03
MFLPPR	— .10	.09
MBFTAP	.06	— .06
MFESDM	.86	— .86
MRABPT	.98	— .99

Tabela 2

OPERATOR PREPOZNAVANJA

	Muško	Žensko
CONST	.7509	.2498
MBKPOP	— .0014	.0013
MAGKUS	— .0050	.0044
MBAP20	— .0014	.0008
MFLPPR	— .0018	.0019
MBFTAP	— .0047	.0049
MFESDM	.0012	— .0009
MRABPT	.0341	— .0360

OUTPUT (DEVICE=PR)

HEADING (TEXT = O N I K E N. T)

TEXT (TEXT = ONIKEN)

*

ONIKEN

* ONIKEN IDENTIFICIRA HIPOTETSKE TAKSONE U JEDNOM SKUPU ENTITETA KOJI JE OPISAN NAD SKUPOM Kvantitativnih varijabli sukcesivnom primjenom jedne modifikacije Q metode taksonomske analize (Momirović, 1969; Rudan, Szivoczka i Momirović, 1978; 1979; Momirović, 1980) i jedne modifikacije generalne Guttmanove procedure za faktorizaciju kvadratnih simetričnih matrica (Stalec i Momirović, 1982) analogne proceduri koja je implementirana u taksonomske programe Zevzek I i Zevzek II nakon toga Oniken određuje sklop i strukturu taksonomskih dimenzija u prostoru standardiziranih varijabli i izračunava njihove korelacije; ovo zbog toga da olakša interpretaciju rezultata.

* Na kraju Oniken proizvodi, implementacijom jednog algoritma za prepoznavanje uzoraka koji je analogan algoritmu implementiranom u program Belzebub (Prot, Bosnar, Stalec i Momirović, 1982), regresijsku matricu koja može biti primijenjena za taksonomsku alokaciju ma kog novog objekta opisanog nad identičnim skupom varijabli nad kojim je opisan i skup objekata na kome je Oniken izveo svoje analize, pod uvjetom da je taj skup bio dovoljno reprezentativan za stabilnu procjenu parametara modela.

* Algoritam i program su opisani u radu Stalec, J. K. Bosnar, F. Prot i K. Momirović Elementarna procedura za identifikaciju kinezioloških taksona. Kineziologija, 12. (1982). 1—2

*

BLOK O. ULAZNE OPERACIJE I REPARAMETRIZACIJA VARIJABLI

*

INPUT (SCORE=BB)

INPUT (SCORE=SS)

CONFORM (IN1=BB, IN2=SS, OUT1=B, OUT2=S)

STATISTICS (SCORE=B, Z=Z)

MULT (A=B, TA, B=B, M=BTB)

DIAGMULT (A=B, D=BTB, C=5, R, M=P)

DELETE (MATRIX=BB)

DELETE (MATRIX=SS)

DELETE (MATRIX=BTB)

*

BLOK 1. MJERE SLIČNOSTI OBJEKATA.

*

HEADING (TEXT=MJERE SLIČNOSTI OBJEKATA, D)

MULT (A=P, B=P, TB, M=PPT)

SCALE (C=PPT, R=C)

PRINT (MATRIX=C, TEXT=MJERE SLIČNOSTI OBJEKATA)

*

BLOK 2 INICIJALNA TAKSONOMSKA SOLUCIJA

*

HEADING (TEXT=INICIJALNA TAKSONOMSKA SOLUCIJA, D)

```
MULT (A=C, B=S, M=CS)
MULT (A=S, TA, B=CS, M=T)
DIAGMULT (A=CS, D=T, C= -0.5, R, M=FF)
SCALE (C=T, R=MM)
INVERSION (R=MM, RINV=MMINV)
MULT (A=FF, B=MMINV, M=AA)
DELETE (MATRIX=CS)
DELETE (MATRIX=FF)
DELETE (MATRIX=MM)
DELETE (MATRIX=MMINV)
DELETE (MATRIX=T)
PRINT (MATRIX=S, TEXT=REALNA ALOKACIJSKA MATRICA)
PRINT (MATRIX=AA, TEXT=INICIJALNA ALOKACIJSKA MATRICA)
```

*

BLOK 3 FINALNA TASONOMSKA SOLUCIJA

*

```
HEADING (TEXT=FINALNA TAKSONOMSKA SOLUCIJA, D)
MULT (A=AA, TA, B=AA, M=AAA)
INVERSION (R=AAA, RINV=AAAINV)
MULT (A=AA, B=AAAINV, M=GAMA)
MULT (A=C, B=GAMA, M=F)
MULT (A=GAMA, TA, B=F, M=T)
DIAGMULT (A=F, D=T, C= -0.5, R, M=ORT)
SCALE (C=T, R=M)
INVERSION (R=M, RINV=MINV)
MULT (A=ORT, B=MINV, M=A)
DELETE (MATRIX=AA)
DELETE (MATRIX=AAA)
DELETE (MATRIX=AAAINV)
DELETE (MATRIX=GAMA)
DELETE (MATRIX=F)
DELETE (MATRIX=T)
DELETE (MATRIX=ORT)
DELETE (MATRIX=MINV)
PRINT (MATRIX=A, TEXT=FINALNA ALOKACIJSKA MATRICA)
PRINT (MATRIX=M, TEXT=KOSINUSI KUTEVA TAKSONOMSKIH VEKTORA, D)
```

*

BLOK 4 SKLOP I STRUKTURA TAKSONOMSKIH DIMENZIJA

*

```
HEADING (TEXT=SKLOP I STRUKTURA TAKSONOMSKIH DIMENZIJA, D)
CROSSCORRELATION (P1=Z, P2=A, R12=F)
CORRELATION (SCORE=A, R=R)
INVERSION
MULT (A=F, B=RINV, M=G)
PRINT (MATRIX=G, TEXT=SKLOP TAKSONOMSKIH DIMENZIJA, D)
PRINT (MATRIX=R, TEXT=KORELACIJE TAKSONOMSKIH DIMENZIJA, D)
PRINT (MATRIX=F, TEXT=STRUKTURA TAKSONOMSKIH DIMENZIJA, D)
DELETE (MATRIX=Z)
DELETE (MATRIX=RINV)
```

*

BLOK 5 PREPOZNAVANJE TAKSONA

*

```
HEADING (TEXT=PREPOZNAVANJE TAKSONA, D)
DIAGONALISATION (R=C)
HOTELLING (NUM=1, F=H)
HADMULT (A=N, TA, CO=OO, M=J)
MERGE (IN1=J, IN2=B, OUT=Y)
DELETE (MATRIX=LAMBDA)
DELETE (MATRIX=X)
DELETE (MATRIX=H)
DELETE (MATRIX=J)
MULT (A=Y, TA B=Y, M=X)
MULT (A=Y, TA, B=A, M=V)
```

```

INVERSION (R=X, RINV=XX)
MULT (A=XX, B=V, M=BETA)
PRINT (MATRIX=BETA, TEXT=OPERATOR PREPOZNAVANJA)
MULT (A=Y, B=BETA, M=ESTAX)
LINEAR (A=S, B=ESTAX, CB= 1.0, M=ERTAX)
PRINT (MATRIX=ESTAX, TEXT=PREPOZNATI TAKSONI)
PRINT (MATRIX=ERTAX, TEXT=POGREŠKE PREPOZNAVANJA)
*
*
*
HEADING (TEXT=KRAJ PROGRAMA ONIKEN)

```

5. LITERATURA

1. Cattell, R. B.: Handbook of multivariate experimental psychology; Rand McNally, Chicago, 1966.
2. Fulgosi, A.: Faktorska analiza; Školska knjiga, Zagreb, 1979.
3. Momirović, K.: Određivanje psiholoških tipova iterativnom primjenom modificirane Q metode faktorske analize; Psihologija, 2, 1, 343—346, 1969.
4. Momirović, K.: Jednostavni algoritam za taksonomsku analizu i prepoznavanje objekata opisanih nad skupom normalnih varijabli. U »Biomedicinska kibernetika«, Skopje, 1980, str. 22—38.
5. Rudan, P., L. Szivoczka i K. Momirović: The application of an algorithm based on Mahalanobis' angles and iterative Q method of taxonomic analysis in the study of micro-evolution; Periodicum Biologorum, 81, 3, 583—589, 1979.
6. Rudan, P., L. Szivoczka i K. Momirović: Analysis of continuous morphological properties in micro-evolution studies: application of an algorithm based on Mahalanobis' angles and iterative Q method of taxonomic analysis; Homo, 30, 4, 252—258, 1979.
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. Štalec, J. i K. Momirović: Jednostavni algoritam za analizu hipotetskih latentnih dimenzija; Kineziologija, 1982, (u štampi).
9. Zakrajšek, E., J. Štalec i K. Momirović: SS-Programski sistem za multivarijantnu analizu podataka; Zbornik simpozija »Kompjuter na Sveučilištu«, Sveučilišni računski centar, Zagreb, 1974, str. c8-1 — c8-16.

THE ELEMENTARY PROCEDURE FOR IDENTIFICATION OF KINESIOLOGIC TAXONS

The algorithm was proposed and program written to identify the hypothetical taxons in a group of entities described over a group of quantitative variables. The algorithm identifies taxons on the basis of similarity among entities, through successive application of a modification of the Q method of taxonomic analysis and a modification of the general Guttman procedure for factorization of square symmetric matrices.

The basic algorithm is accompanied by 2 modules. The first one determines the system and structure of taxonomic dimensions in the space of standardized variables and the second produces, under the general linear model, the operator of recognition that can be applied in taxonomic allocation of any new object described over an identical group of variables.

Янез Шталец, Ксения Боснар, Франье Прот, Константин Момирович

ЭЛЕМЕНТАРНАЯ ПРОЦЕДУРА ДЛЯ ИДЕНТИФИКАЦИИ КИНЕЗИОЛОГИЧЕСКИХ ТАКСОНОВ

Предложен алгоритм и написана программа, которые идентифицируют гипотетические таксоны в одном множестве заданий, которое описано над одним множеством переменных. Алгоритм идентифицирует таксоны на основании сходства заданий при помощи последовательного применения Q-метода таксономического анализа и модификации генеральной процедуры Гутмана для факторизации квадратных симметрических матриц.

К основному алгоритму присоединены два модуля. Первый из них определяет комплекс и структуру таксономических измерений в пространстве стандартизованных переменных, а второй производит, при помощи генеральной линейной модели, оператор узнавания, который может быть применен для таксономической алокации любого нового объекта, описанного над таким же множеством переменных.

