

KONSTANTIN MOMIROVIĆ I ŽIVAN KARAMAN
Sveučilišni računski centar, Zagreb

Primljeno 7. 4. 1982.

**GOLDDIFF — ALGORITAM I PROGRAM ZA ANALIZU
KVANTITATIVNIH PROMJENA SPRKTRALNOM
DEKOMOZICIJOM UNIVARIJATNIH TRAJEKTORIJA**

P. Karaman
P. Momirović

SAŽETAK

Konstruirani su model i algoritam i napisan je program za spektralnu analizu univarijantnih promjena stanja grupe objekata. Trajektorije promjena dekomponiraju se na ortogonalne komponente da bi se time pojednostavnio prikaz promatranih stohastičkih procesa, pa se zatim pomoću tih zadržanih komponenata reproduciraju originalne trajektorije promjena.

1. UVOD

U ovom radu opisani su model, algoritam i program koji su namijenjeni spektralnoj analizi jednodimenzionalnih promjena stanja neke grupe objekata na kojima je kroz neki vremenski period registrirana promjena jedne kvantitativne varijable. Takav tip analize potreban je u kineziologiji, ekonomiji, tehnologiji i svim onim područjima znanosti gdje se višekratno prate promjene neke varijable. Tako, napr., u kineziologiji to može biti grupa ljudi kojima se kroz deset minuta mjeri broj okretaja na sobnom biciklu, u ekonomiji to mogu biti mjesečne promjene proizvodnje nekog važnog artikla promatrane u nizu zemalja, a u tehnologiji to može biti promjena temperature registrirana prilikom nekoliko ciklusa nekog tehnološkog procesa.

Modelom koji je ovdje predložen mogu se analizirati ne samo nestacionarni procesi, nego i procesi koji imaju striktno stacionarna stanja, jer model dopušta singularitete. Spektralnom dekompozicijom trajektorija promjena reduciramo ih na ortogonalne komponente promjena koje su opisane kako u prostoru vremena tako i u prostoru entiteta. Pomoću tih ortogonalnih komponenata možemo reproducirati originalne trajektorije promjena metodom Eckarta i Younga (1936.).

Ovaj model je originalno predložio L. Tucker, a potpun opis izvornog modela i algoritma nalazi se u Momirović (1972.). Ovdje je detaljno opisan algoritam koji je implementiran u programu COLDDIFF, napisanom u meta-jeziku SS (Zakrajšek, Štalec, Momirović, 1974.) i implementiranom na računaru UNIVAC 1100 Sveučilišnog računskog centra u Zagrebu.

2. MODEL

Pretpostavimo da smo registrirali promjene neke kvantitativne varijable u n vremenskih točaka na skupu od m entiteta, i na taj način dobili matricu trajektorije promjena B , tipa $m \times n$. Pri tome moraju biti zadovoljeni slijedeći uvjeti:

- $m \geq n$
- n vremenskih točaka je ekvidistantno.

Iako se eksplicitno pretpostavlja samo da je $m \geq n$, razumni se rezultati mogu očekivati samo ako je $m \geq 3 * n$.

Spektralna analiza trajektorije promjena izvrši se tako da se nađu svojstvene vrijednosti matrice $B^T B$, a značajnim vrijednostima spektra trajektorije promjena proglašene su one svojstvene vrijednosti koje su veće od prosjeka svih svojstvenih vrijednosti. Faktori entiteta definiraju se kao zadržane glavne komponente, a faktori mjerenja su definirani kao glavne osi matrice $B^T B$.

Trajektorija promjena aproksimira se Eckart-Youngovom metodom aproksimacije matrice drugom matricom manjeg ranga, a točnost aproksimacije mjeri se kvocijentom tragova aproksimirane i originalne matrice podataka.

3. ALGORITAM

3.1. Parametri i distribucije mjerenja

Algoritam i program COLDDIFF analiziraju podatke u izvornoj matrici. Ukoliko se podatke želi na neki način reparametrizirati, vjerojatno je najrazumniji način da se to uradi tako da se podaci centriraju na aritmetičku sredinu i standardiziraju na standardnu devijaciju prve (ili posljednje) vremenske točke.

Parametre mjerenja dobijemo tako da za svaku vremensku točku izračunamo očekivanje, varijancu, frekvencije po razredima i sve ostale parametre kako je to inače uobičajeno za varijable (svaka vremenska točka predstavljena je jednim stupcem u matrici B). Ako sa $\underline{1}^T = (1, 1, \dots, 1)$ označimo m -dimenzionalni sumacioni vektor, tada je matrica kovarijanci vremenskih točaka dana sa

$$\underline{C} = (B^T B - B^T \underline{1} B) * 1/m$$

gdje je $\underline{J} = \underline{1} \underline{1}^T * 1/m$. Matrica varijanci vremenskih točaka dana je sa

$$\underline{S} = \text{diag } \underline{C}$$

pa je matrica korelacija vremenskih točaka

$$\underline{R} = \underline{S}^{-1/2} \underline{C} \underline{S}^{-1/2}$$

3.2. Spektralna dekompozicija matrice podataka

Da bismo dobili komponente promjena, izračunamo skalarnu produktove vektora vremenskih točaka

$$\underline{D} = \underline{B}^T \underline{B},$$

te izračunamo svojstvene vrijednosti i vektore matrice \underline{D} . Matricu \underline{D} sada možemo prikazati u obliku

$$\underline{D} = \underline{X} * \underline{E} * \underline{X}^T$$

ako sa \underline{X} označimo matricu svojstvenih vektora, a sa \underline{E} dijagonalnu matricu svojstvenih vrijednosti matrice \underline{D} .

Kako su sve svojstvene vrijednosti matrice \underline{D} negativne, matricu \underline{B} možemo prikazati u obliku

$$\underline{B} = \underline{Y} * \underline{L} * \underline{X}^T,$$

ako sa \underline{L} označimo dijagonalnu matricu drugih korijena svojstvenih vrijednosti matrice \underline{D} .

$$\underline{L} = \underline{E} ** (0.5)$$

a sa \underline{Y} označimo desne svojstvene vektore matrice \underline{B} ,

$$\underline{Y} = \underline{B} * \underline{X} * \underline{L} ** (-1).$$

Budući ne želimo zadržati potpun skup svojstvenih vrijednosti, reducirati ćemo ih na taj način da zadržimo samo one svojstvene vrijednosti koje su veće od prosjeka svih svojstvenih vrijednosti,

$$\lambda > = \sum_{i=1}^n \lambda_i / n,$$

Broj zadržanih svojstvenih vrijednosti označimo sa k . Zadržane svojstvene vrijednosti su sigurno striktno pozitivne, bez obzira na to da li je matrica \underline{D} pozitivno definitna ili samo pozitivno semidefinitna. Sa zadržanim svojstvenim vrijednostima možemo reducirati originalnu matricu trajektorija promjena pomoću izraza

$$\underline{B} = \underline{\tilde{Y}} * \underline{\tilde{L}} * \underline{\tilde{X}},$$

ako smo sa $\underline{\tilde{L}}$ označili dijagonalnu matricu zadržanih svojstvenih vrijednosti (tipa $k \times k$), a $\underline{\tilde{X}}$ i $\underline{\tilde{Y}}$ su pripadne reducirane matrice svojstvenih vektora.

3.3. Faktori entiteta i faktori mjerenja

Faktori entiteta definirani su kao nestandardizirane zadržane glavne komponente,

$$\underline{G} = \underline{B} * \underline{\tilde{X}},$$

dok su faktori mjerenja definirani kao glavne osovine matrice \underline{D} ,

$$\underline{H} = \underline{\tilde{X}} * \underline{\tilde{L}} ** 0.5.$$

3.4. Aproksimacija matrice podataka

Originalnu matricu trajektorija promjena \underline{B} aproksimiramo metodom Eckarta i Younga (1936.), na ovaj način:

$$\underline{\tilde{B}} = \underline{B} * \underline{\tilde{X}} * \underline{\tilde{X}}^T$$

i to je pod danim modelom najbolja (po kriteriju najmanjih kvadrata) aproksimacija matrice \underline{B} matricom ranga $k < n$.

Ako rezidualnu matricu podataka označimo sa $\underline{\hat{B}} = \underline{B} - \underline{\tilde{B}}$, točnost te aproksimacije možemo mjeriti sa

$$\text{trag}(\underline{\hat{B}}^T \underline{\hat{B}}) / \text{trag}(\underline{B}^T \underline{B}) = \beta,$$

a taj koeficijent β je upravo jednak postotku koji čini suma nezadržanih svojstvenih vrijednosti u sumi svih svojstvenih vrijednosti matrice $\underline{B}^T \underline{B}$.

4. PROGRAM

Program COLDIFF napisan je u meta jeziku SS i implementiran na računalu UNIVAC 1100 Sveučilišnog računskog centra u Zagrebu. U trenutnoj verziji (5.2/M) sistema SS njime je moguće analizirati promjene stanja ne više od 10000 entiteta promatranih u ne više od 250 vremenskih točaka. Podacima je potrebno pridružiti naredbe koje opisuju format podataka u skladu s uputama za sistem SS (jednu SEQUENCE naredbu i onoliko VARIABLE naredbi koliko ima vremenskih točaka).

COLDIFF analizira matricu podataka u izvornoj matrici. Ako je zbog nekog razloga potrebno reparametrizirati podatke, vjerojatno je najrazumnije da ih se centriraju na vrijednost prosjeka prvog mjerenja i skaliraju na vrijednost standardne devijacije prvog mjerenja. Ovo korisnik mora uraditi sam i pridružiti COLDIFFu tako, ili na drugi način, reparametrizirane podatke. Postoji verzija programa COLDIFFC koja sama na gore navedeni način reparametrizira podatke.

5. LITERATURA

1. Eckart, C. T. & Young, G.: THE APPROXIMATION OF ONE MATRIX BY ANOTHER OF LOWER RANK. *Psychometrika*, 1, 211—218, 1936.
2. Momirović, K.: METODE ZA TRANSFORMACIJU I KONDENZACIJU KINEZILOŠKIH INFORMACIJA. Institut za kineziologiju, Zagreb, str. 250—254, 1972.
3. Zakrajšek, E., Štalec, J., Momirović, K.: SS — PROGRAMSKI SISTEM ZA MULTIVARIJATNU ANALIZU PODATAKA. I. Međunarodni simpozij »Kompiuter na sveučilištu«, Zagreb, 1974.

COLDIFF — THE ALGORITHM AND PROGRAM FOR ANALYSIS OF QUANTITATIVE CHANGES BY MEANS OF THE SPECTRAL DECOMPOSITION OF UNIVARIANT TRAJECTORIES

The model and algorithm were constructed and the program was written for the spectral analysis of univariant changes in condition of a group of objects. Trajectories of changes were decomposed into orthogonal components in order to simplify the representation of the observed stochastic processes, in order to reproduce, by means of these retained components, the original trajectories of changes.

Константин Момирович, Живан Караман

»КОЛДИФ« — АЛГОРИТМ И ПРОГРАММА ДЛЯ АНАЛИЗА КОЛИЧЕСТВАННЫХ ИЗМЕНЕНИЙ ПРИ ПОМОЩИ СПЕКТРАЛЬНОГО РАСЧЛЕНЕНИЯ УНИВАРИАНТНЫХ ТРАЕКТОРИЙ

Созданы модель и алгоритм и написана программа для спектрального анализа унивариантных изменений состояния группы объектов. Траектории изменений расчленяются на ортогональные компоненты с целью упрощения изображения разматриваемых стохастических процессов, а затем при помощи этих сохраненных компонентов создаются оригинальные траектории изменений.