

FŐKOMPONENSANALIZIS MODELLEK ÖSSZEHASONLÍTÁSA KIVÁLTOTT POTENCIÁL
ADATOKON

Czobor Pál, Vitrai József, Simon Gábor, Ivanyos Gábor, Varga László,
Marosfi Sándor

Semmelweis OTE Pszichiátriai Klinika, Semmelweis OTE Számítástechnikai
Csoport, ELTE Matematikus szak, V. évfolyam, Budapesti Hőerőmű Vállalat
Számítástechnikai Csoport

Annak ellenére, hogy a lineáris modell érvényessége a kiváltott potenciálok /KP/ esetében elvileg erősen vitatható, a gyakorlatban /a KP-ok adatfeldolgozásának gyakorlatában/ főkomponens-analízis /PCA/ alkalmazásával jó eredmények érhetők el: a legtöbb szerző egyetért abban, hogy mindössze néhány /2-6/ faktor segítségével már igen jó adatredukció lehetséges /a megmagyarázott variancia értéke általában 70-95 %/ [1, 2, 3]. A módszerben rejlő ilyen lehetőség más adattömörítő módszerekkel történő összehasonlítás után még inkább szembetűnő [4].

A fenti eredmények ellenére a PCA módszerrel kapcsolatban ugyanakkor számos módszertani probléma vár megoldásra. Ezeknek a problémáknak a vonatkozásában gyakran a priori szakmai ismereteink nincsenek. Így igen gyakran nehézséget okozhat a megfelelő PCA modell kiválasztása, hiszen a "kitűzött cél és az alkalmazott módszer egymással interakcióban áll" [5].

Vizsgálatainkban különböző vizuális ingerekre kapott KP-okra elvégzett - kovariancia /cPCA/ továbbá korrelációs mátrixból /rPCA/ kiinduló varimax rotáció nélküli valamint varimax /PCAv/ rotációt is magukba foglaló - PCA-k eredményeit kívántuk összehasonlítani. Az összehasonlítás szempontjai: 1./ adatredukció önmagában, 2./ adatredukciós és egyidejűleg figyelembe vett osztályozási kritérium.

Módszer

A vizsgálatokban 8 egészséges fiatal személy /életkor: 20-28 év/ vett részt. A vizsgálati sorozat délelőtt 9 óra körül indult és másnap 8 óra körül fejeződött be. Ez idő alatt kb. 90 percenként 4 különböző vizuális inger /flash = F, saktábla = CH, piktogram = A, piktogram inger, amelyet egy másik piktogram "fed" = MA/ segítségével átlagolt kiváltott válaszokat vettünk fel /elvezetés: $O_z - A_{12}$; $C_z - A_{12}$ /, 3 személynél a F ingert minden ülés végén megismételtük. Egy-egy 500 ms-os potenciált 100 mintavételi pont reprezentált.

Minden egyes személy KP-jait külön-külön többféle PCA-nak vetettük alá, majd az eredményül kapott PC-értékek alapján a különféle ingerekre kapott KP-okat SWDA segítségével próbáltuk elkülöníteni. Egy-egy személyről 8 PCA és 8 SWDA futást nyertünk, a két elektróda pozíciót is figyelembe véve.

A programok a BMDP 1977-es verziójának 4M ill. 6M programjai voltak. Az SWDA-ban a terminálási kritérium a Wilks-lambda csökkenése /discriminant method:2/, az F-to-enter 3.01, az F-to remove 3.0 volt. Klasszifikációra a jackknife módszert használtuk.

Eredmények

Kovariancia vs. korreláció

A cPCA és rPCA-k kumulatív százalékaiknak különbségeit az összes személyre nézve átlagolva a kapott átlagos különbség minden összehasonlításban és minden PC sorszámánál pozitív előjelű /I. táblázat/. A táblázatból az is megfigyelhető, hogy a két módszer ilyen szempontból vett különbsége a bevont PC-k számának növekedésével csak viszonylag "lassan" csökken.

I. táblázat

A cPCA és rPCA által megmagyarázott variancia százalékok közötti átlagos különbség / S.D.

a PC-k sorszáma	occipitalis		centralis	
	nem rotált	rotált	nem rotált	rotált
1.	12 ₁₁	15 ₁₂	2 ^a ₄	7 ₆
2	11 ₅	17 ₇	7 ₄	9 ₅
4	8 ₃	16 ₆	7 ₃	12 ₆
6	5 ₂	11 ₅	4 ₂	10 ₅

a: nem szignifikáns

Ugyanazon adatokon alapuló cPCA-k és rPCA-k eredményei általában különböznek. Megegyezés csupán abban a speciális esetben várható, ha a szórások pontonként megegyeznek. Annak ellenére azonban, hogy a szórások az elemzett görbéken pontonként eltérnek, megállapítható, hogy a cPCA első PC-je az rPCA első két PC-je közül valamelyiknek viszonylag jól megfelel / $r > 0,8$ /. A kétféle PCA-ból nyert loading függvények közötti hasonlóság mértékeként a korrelációs koefficiens alapul véve - a maximális korrelációk megkeresése után - megállapítható: a cPCA során kapott PC 5 személynél az rPCA során is első PC lesz, míg 3 személy esetén 2. PC lesz.

Adatredukció: a rotálás hatása

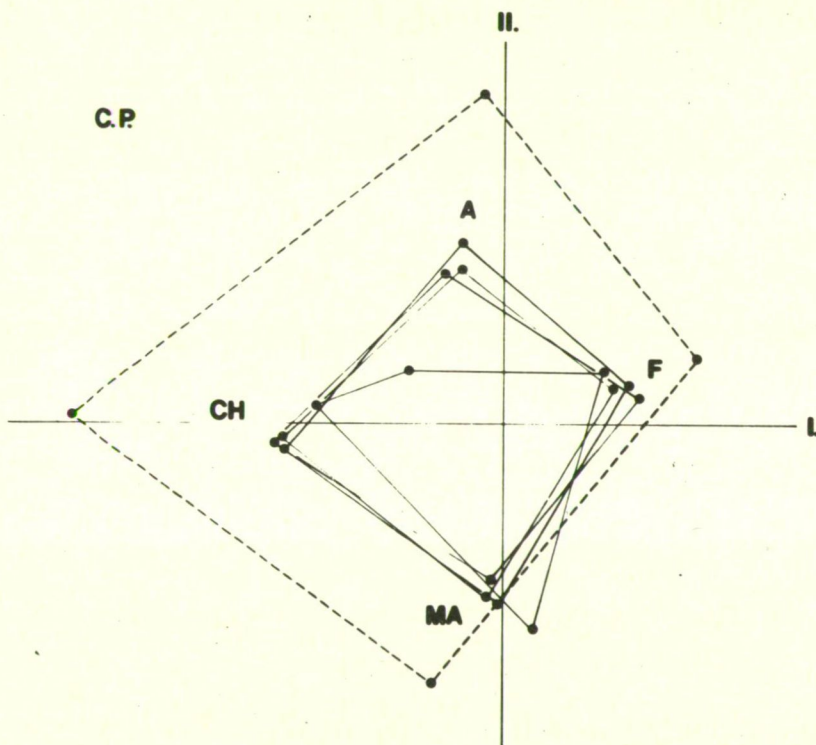
Adataink szerint az első 10 rotálatlan PC által létrehozott alter /mind a kovariancia, mind a korrelációs mátrix esetén/ átlagosan a KP-ok összvarianciájának 96 %-át leírja és a megmagyarázott variancia minden esetben 95 % fölött van. Ezeknek az altereknek az új bázisait kerestük meg a varimax rotáció kritériumát felhasználva. A különböző futások összegzett adatai azt mutatják, hogy az egyes rotált PC-k által megmagyarázott variancia a legvalószínűbben a 2. és 3. PC

után haladja meg az ugyanilyen sorszámú rotálatlan PC varianciáját.

A rotálatlan és rotált PC-k által megmagyarázott variancia kumulatív százaléka között az első két PC után az eltérés közel 20 %-os, de még 6 PC után is jelentős különbség tapasztalható a rotálatlan változat javára.

Osztályozás

Ha a négy különböző PCA módszer után az elválasztásba bevont összes PC-ből kanonikus változókat képezünk és a legjobban elválasztó két kanonikus változó terében ábrázoljuk a képekre kapott csoportátlagokat, feltűnő, hogy a különböző módszerekkel kapott eredmények között lényeges eltérés nincs /1. ábra/: a kanonikus változók terében a



1. ábra. A különböző képekre kapott kiváltott válaszok elhelyezkedése a két legjobban elválasztó kanonikus változó terében. Szaggatott vonal = eredeti adatokból képezett kanonikus változók tere; folytonos vonal = a négy különböző PCA után kapott kanonikus változók tere /C.P. = vizsgált személy/.

négy képre kapott EP-k átlagai által meghatározott alakzatok egymást gyakorlatilag teljes mértékben lefedik. Ugyanakkor ezek az alakzatok az eredeti EP változókból - hasonló módon - kapott alakzatokkal is nagymértékű hasonlóságot mutatnak, jelezve, hogy a képek közötti intraindividuális viszonyok csak kissé torzultak. A PC-k-ből kiindulva azonban a képek távolsága általában kisebb, hiszen a módszer szétválasztási és adatredukciós célokat együttesen figyelembe véve

nem optimális.

Ha az adatredukció és szétválasztás követelményét a különböző PCA + SWDA eljárások összehasonlítása során fokozottan figyelembe akarjuk venni /esetleg 10-nél jóval kevesebb PC-t akarunk használni/, akkor a legjobb jackknife százalékok önmagukban kevésbé informatívak, hiszen számításba kell vennünk az elválasztáshoz szükséges kiszámítandó PC-k számát is. Ennek figyelembevételére egy olyan paramétert alakítottunk ki, amelyben a szétválasztási százalékot a legnagyobb indexű PC sorszámával osztottuk. Így lényegében egy olyan számhoz jutunk, amely egy PC szétválasztásra gyakorolt átlagos hatását mutatja a legjobb szétválasztás esetén. Az utóbbi paraméter vonatkozásában a cPCA és rPCA változatok között occipitálisan és centrálisan is átlagosan kb 4 % eltérés van az elválasztásban szereplő egyes PC-k "teljesítménye" között. Ez az utóbbi különbség adatredukciós szempontból már jelentősnek mondható, hiszen figyelembe véve, hogy a besorolásnál átlagosan több mint 4 PC bevonódik, a különbség négy PC után már kb 16 % lesz.

Megbeszélés

A cPCA és rPCA módszer használata esetén lényegében egyetlen jelenséget tükröző különböző folyamatok optimális közelítése a cél. Azt, hogy a sztochasztikus kapcsolat mértékeként a kovarianciát, vagy a korrelációt használjuk-e, vagyis, hogy "a standardizálás kívánatos-e a végső analízisben, nem statisztikai alapokon kell eldönteni" [6]. Abban az esetben, ha a vektorváltozó komponensei azonos mértékűek - ami a KP-ok esetében is fennáll - nincs a priori ok a korrelációs mérték választására [7].

Annak ellenére, hogy az eredeti - és a standardizált KP mérték-tér dimenzionalitása elméletileg azonos, a kétféle tér effektív dimenzionalitása /vagy "domináns dimenzionalitása", Rao [8]/ adataink szerint különbözik. Ez abból következik, hogy a cPC-k esetén a megmagyarázott variancia kumulatív százalékának növekedése - az adatredukció szempontjából leginkább fontos első PC-kre nézve - jóval gyorsabb, mint az rPC-k esetén.

Hasonló eredményekre jutott vizsgálatait során John [9] is. Véleménye szerint, abban az esetben, ha kovariancia mátrixot használ, akkor a tér "apparent dimenzionalitása" kisebb lesz: "Mivel nem minden jel egyforma teljesítményű, azt várjuk, hogy a kapott eredmény viszonylag egyszerűbb lesz, mivel az erősebb jelek hatását fogja tükrözni".

Ugy gondoljuk az előbbi magyarázatot azzal lehetne kiegészíteni, hogy az adatleírás szempontjából nem szerencsés, hogy a korrelációs mátrixban az alacsonyabb - relative több zajösszetevőt tartalmazó - amplitudókat a kevésbé zajos jelamplitudókkal azonos módon súlyozzuk. Ilyenkor ugyanis nemcsak a leírandó hullámforma változatok száma növekedhet, hanem a jelek közötti összefüggésrendszer elemezhetősége is kevésbé megbízhatóvá válhat.

A különböző módszerek PC-iből képzett kanonikus változók terében az egyes képekre kapott KP-ok átlagai hasonló módon helyezkednek el. Ez az elhelyezkedés hasonló a képeknek az eredeti adatokból képzett kanonikus változók terében való elhelyezkedéséhez. Ha tehát az adatredukció követelményét nem tartjuk szigorúan szem előtt, és az adatokat egy átlagosan 96 % varianciát magyarázó altérbe vetítjük a PC-k segítségével, a képek intraindividuais KP-térben elfoglalt hely-

zetét viszonylag torzítás nélkül becsülhetjük. Meg kell azonban je-gyezni, hogy mivel a PCA + SWDA adatredukciós és szétválasztási szem-pontot egyszerre figyelembe véve nem optimális, még ilyen jelentős variancia megtartás mellett is a képek közötti távolságok az eredeti változók teréből képezett kanonikus változók terében nagyobbak, to-vábbá a szétválasztás százalécai is jelentősen javulnak.

A PCA utáni adatfeldolgozás - a jelentős adatredukció ellenére is - számos problémát rejt magában. Mivel utófeldolgozási célokra igen gyakran egyváltozós módszereket /varianciaanalízis, t-teszt/ használnak, a variancia alapján jelentősnek tűnő PC változók viszony-lag nagy száma miatt még mindig meglehetősen sok számolás szükséges. Emiatt a kutatók gyakran a PC változók preszelekciójára kényszerülnek és az összehasonlításokat csak az első néhány PC vonatkozásában végzik el. Adataink /melyeket az elválasztás százaléka és az elválasztásban szereplő PC-k indexeinek együttes figyelembevételével kaptunk/ arra hívják fel a figyelmet, hogy egy ilyen preszelekció az irodalomban leggyakrabban használt rPCAV módszer esetén - összehasonlítva a cPCA módszerrel - jelentős információvesztést okozhat. /Például a centrális elvezetésben az rPCAV után a diszkriminálásban legjelen-tősebb, SWDA-ba elsőként bevont PC sorszáma 4 esetben volt négynél na-gyobb. Az ezen PC-k által megmagyarázott variancia egyetlen PC-nél sem haladta meg a 10 %-ot./

Irodalomjegyzék

- [1] Perry, N.W. and Childers, D.G.: The human visual evoked response. Method and theory. Charles C. Thomas, Springfield, 1969.
- [2] John, E.R.: Neurometrics: Clinical Applications of Quantitative electrophysiology. In: E.R. John and R.W. Thatcher /Eds/: Functional Neuroscience /Vol. 2/ Lawrence Erlbaum Associate Publishers, Hillsdale, New Jersey, 1977.
- [3] Kavanagh, R.N., Terrance, M.D. and Fender, D.H.: The dimension-ality of the human evoked potential. Electroencephalogr. Clin. Neurophysiol., 1975, 40: 633-644.
- [4] John, E.R. and Ruchkin, D.S.: Signal analysis of evoked poten-tials recorded from cats during conditioning. Science, 1963, 141: 429-431.
- [5] Donchin, E. and Heffley, E.: Multivariate analysis of event-related potential data: A tutorial review. In: D.A. Otto /Ed./: Multidisciplinary perspectives in event-related brain potential research. EPA-600/9-77-043, U.S. Government Printing Office, Washington, D.C. 1978, 555-572.
- [6] Kendall, M.G. and Stuart, A.: The advanced theory of statistics. Vol. 3. Design and Analysis, and Time Series, Griffin, London, 1976. p. 296.
- [7] Gnanadesikan, R. and Wilk, M.B.: Data analytic methods in multivariate statistical analysis. In: P.R. Krishnaiah /Ed./: Multivariate analysis - II. Academic Press, New York and London, 1969: 593-638.
- [8] Rao, C.R.: Linear statistical inference and its applications. John Willey, New York, 1965.
- [9] John, E.R., Ruchkin, D.S. and Villegas, J.: Experimental background: signal analysis and behavioral correlates of evoked potential configurations in cats. Annals of New York Academy of Sciences, 1964, 112: 362-420.