Originalni naučni rad UDK: 681.513.8 : 007.52 : 528.854
BIBLID:0350-8528(2005),17.p. 19-29

VREMENSKI ORIJENTISAN HIJERARHIJSKI METOD ZA IZRAČUNAVANJE PCA I MCA

Marko Janković
Elektrotehnički institut "Nikola Tesla", Beograd

Sadržaj: *U radu je opisan prvi opšti metod koji omogućava transformaciju neuralnih mreža za izračunavanje glavnog/sporednog potprostora (PSA/MSA) u neuralne mreže za izračunavanje glavnih/sporednih komponenata (PCA/MCA). Primenom ovog metoda se mogu dobiti neuralne mreže koje su pogodne za realizaciju u paralelnom hardveru jer je moguće ostvariti mreže (algoritme) koje su potpuno homogene sa aspekta pojedinačnog neurona.*

Ključne reči*: analiza glavnih komponenata, analiza sporednih komponenata, analiza glavnog potprostora, analiza sporednog potprostora, vremenski orijentisan hijerarhijski metod, neuralne mreže.*

1. UVOD

Analiza glavnih komponenata (PCA) je jedna od najstarijih i najpoznatijih metoda za multivarijabilnu analizu i data mining (pronalaženje podataka u veoma velikim bazama). PCA je prvi put predstavljena od strane Pirsona [1] koji ju je koristio u biološkom kontekstu, i kasnije razvijana od strane Hotelinga [2] koji je obavljao istraživanja na području psihometrije. PCA je nezavisno razvijena od strane Karhunena u kontekstu teorije verovatnoće i kasnije generalizovana od strane Leva [3]. Poslednjih godina, razvijeni su mnogobrojni efikasni adaptivni algoritmi za implementaciju PCA, analize sporednih komponenata (MCA), kao i njihovih ekstenzija [4-8].

Generalno govoreći, namena PCA je određivanje relativno malog broja glavnih komponenata (međusobno nekorelisanih) koje reprezentuju skup slučajnih signala sa srednjom vrednošću nula na ulazu sistema i koje pri tome zadržavaju maksimalnu moguću količinu informacija koja se sadrži u originalnim signalima.

Među ciljevima PCA su:

1. smanjenje dimenzije signala,
2. otkrivanje linearne zavisnosti među komponentama signala,
3. određivanje karakterističnih osobina signala,
4. vizuelizacija višedimenzionih podataka,
5. identifikacija grupa objekata i tako dalje.

Takođe, PCA je korisna kao pretprocesirajući korak za analizu nezavisnih komponenata (engleski: Independent Component Analysis – ICA).

PCA je široko korišćena i proučavana u oblasti prepoznavanja oblika i procesiranja signala. To je veoma važan alat u mnogim inženjerskim i naučnim disciplinama kao što su, na primer, kompresija podataka, otkrivanje karakterističnih osobina signala, filtriranje šuma, restauriranje signala, klasifikacija [9]. PCA se koristi u data mining kao tehnika za redukciju podataka. U oblasti obrade slike i kompjuterskog vida PCA reprezentacije su korišćene za rešavanje problema kao što su: prepoznavanje lica i objekata, praćenje pokretnog objekta, detekciju, parametrizaciju oblika, pojava objekta ili kretanje objekta [10, 11].

Često su glavne komponente (pravci u kojima ulazni podaci imaju najveće varijanse) smatrani za važne dok su komponente koje su vezane za sporedne komponente smatrane nevažnim ili povezivane sa šumom. U poslednje vreme je pokazano, da MCA (minor component analysis) predstavlja veoma koristan alat za procesiranje sekvenci podataka. MCA je korisna na primer kod: praćenja pokretnih meta [12], procenjivanja pozicije emitera signala [13] kao i procene parametara signala [14], analize i razumevanja bioloških podataka [15], potiskivanja šuma ili aproksimacije funkcija pomoću fitovanja [16].

Uopšteno govoreći, PCA je povezana i motivisana sa dva sledeća problema:

1. Za date slučajne vektore $x(k) \in \Re^K$, koji imaju konačne druge momente i srednju vrednost jednaku nuli, potrebno je naći linearni potprostor manje dimenzije N $(N < K)$ za koji je rastojanje x od tog potprostora minimalno. Ovaj problem se pojavljuje kod kompresije podataka gde je zadatak reprezentovanje svih podataka sa redukovanim brojem parametara pri čemu je potrebno obezbediti minimalnu distorziju usled projekcije.
2. Za dati slučajan vektor $x(k) \in \Re^K$, nađi N-dimenzioni linearni potprostor koji sadrži većinu varijanse podataka x. Ovaj problem je vezan za ekstrakciju karakterističnih osobina signala, gde je cilj redukcija dimenzije podataka pri čemu je potrebno zadržati maksimalnu količinu informacija koja se sadrži u ulaznim podacima.

Pokazalo se da oba problema imaju isto optimalno rešenje (u smislu najmanje kvadratne greške) koje se bazira na statistici drugog reda, posebno, na sopstvenim vektorima kovarijansne matrice ulaznog signala x i to je rešenje u suštini ekvivalentno Karhunen-Levovoj transformaciji koja se koristi kod analize slike i procesiranja signala. Drugim rečima, PCA je tehnika za izračunavanje sopstvenih vrednosti i sopstvenih vektora estimirane kovarijansne matrice

$$C = \mathrm{E}\left\{x(k)x(k)^{\mathrm{T}}\right\} = V\Lambda V^{\mathrm{T}} \in \Re^{KxK} \tag{1}$$

gde je $\Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_K\}$ dijagonalana matrica koja sadrži K sopstvenih vrednosti i $V=[v_1, v_2, \ldots, v_K] \in \Re^{KxK}$ je odgovarajuća ortonormalna ili unitarna matrica koja se sastoji od sopstvenih vektora jedinične dužine.

Karhunen-Levova transformacija određuje linearnu transformaciju ulaznog vektora x kao

$$y_p = V_S^{\mathrm{T}} x \tag{2}$$

gde je $x=[x_1(k), x_2(k), \ldots, x_K(k)]^{\mathrm{T}}$ ulazni vektor sa srednjom vrednošću nula, $y_p=(y_1(k), y_2(k), \ldots, y_N(k))^{\mathrm{T}}$ izlazni vektor koji se naziva vektorom glavnih komponenti i $V_s = [v_1, v_2, \ldots, v_N]^{\mathrm{T}} \in \Re^{KxN}$ predstavlja sopstvene vektore signalnog potprostora, gde su vi = $[v_{i1}, v_{i2}, \ldots, v_{iK}]^{\mathrm{T}}$ ortonormalni vektori . Vektori v_i $(i = 1,2,\ldots,N)$ su sopstveni vektori kovarijansne matrice, dok su varijanse glavnih komponenata y_i odgovarajuće sopstvene vrednosti. Sa druge strane K-N sporednih komponenata je dato sa

$$y_M = V_N^{\mathrm{T}} x, \tag{3}$$

gde V_N [v_K, v_{K-1}, ..., v_{K-N+1}] sadrži (K-N) sopstvenih vektora povezanih sa najamanjim sopstvenim vrednostima.

Tako, osnovni problem koji pokušavamo da rešimo predstavlja standardni problem izračunavanja sopstvenih vektora-vrednosti koji se može formulisati sledećim skupom jednačina

$$Cv_i = \lambda_i v_i \quad (i = 1,2,\ldots,N) \tag{4}$$

gde su v_i sopstveni vektori, a λ_i odgovarajuće sopstvene vrednosti i $C = E(xx^T)$ kovarijansna matrica ulaznog signala x, čija je srednja vrednost nula i E je operator očekivanja. Zapazimo da se jednačina (4) može pisati i u formi $V^TCV=\Lambda$, gde je Λ dijagonalna matrica sopstvenih vrednosti matrice C.

U standardnim numeričkim pristupima određivanja glavnih komponenata, prvo se mora izračunati kovarijansna matrica C i onda se pristupa određivanju sopstvenih vektora i odgovarajućih sopstvenih vrednosti pomoću nekog od poznatih algoritama. Međutim, u situacijama kada je ulazni vektor velikih dimenzija (na primer 1000 elemenata), kovarijansna matrica postaje veoma velikih dimenzija (10^6 ulaza) i tada izračunavanje sopstvenih vektora može postati veoma teško.

Upotrebom neuralnih mreža sa odgovarajućim adaptivnim zakonima učenja omogućava se izračunavanje sopstvenih vektora i njima pridruženim sopstvenih vrednosti na osnovu ulaznog signala $x(k)$ bez potrebe za izračunavanjem ili estimiranjem veoma velike matrice C. Takav pristup je posebno koristan u slučajevima kada je ulazni signal nestacionaran, na primer u slučajevima kada se zahteva praćenje sporih promena međuzavisnosti ulaznih podataka (signala) ili pri osvežavanju sopstvenih vektora novim odbircima. Izračunavanje kovarijansne matrice samo za sebe je računarski veoma skupo. Dalje, direktna dijagonalizacija matrice ili *eigenvalue decomposition* mogu takođe biti veoma računarski skupe jer je broj operacija koji je u tom slučaju potrebno izvesi $O(K^3)$. Većina adaptivnih algoritama koje ćemo sada predstaviti ne zahteva izračunavanje kovarijansne matrice i ima malu složenost.

2. VREMENSKI ORIJENTISAN HIJERARHIJSKI METOD

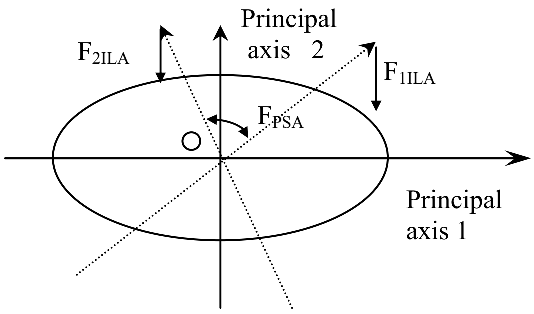
U ovoj sekciji će biti predstavljen opšti metod koji transformiše PSA/MSA algoritme u PCA/MCA algoritme. Glavna ideja koja omogućava primenu metoda je data u sledećoj rečenici:

Svaki neuron radi ono što je najbolje za njegovu "porodicu" a potom ono što je najbolje za njega samog.

Ovu ideju ćemo nazvati "porodični princip". Drugim rečima algoritam se sastoji od dva dela: jedan deo je odgovoran za ostvarivanje porodičnog cilja *(family-desirable feature learning)* a drugi je odgovoran za ostvarivanje želja pojedinačnih neurona (individual-neuron-desirable feature learning). Drugi deo zakona učenja se uzima sa težinskim koeficijentom α koji je po apsolutnoj vrednosti manji od 1. To drugim rečima znači da se u algoritam uvodi vremenski orijentisana hijerarhija kod realizacije "porodičnog" i "individualnog" dela zakona učenja.

Na slici 1 je prikazana gustina verovatnoće multivarijabilnog Gausovog signala. Poznato je da su u tom slučaju glavne ose hiperelipsoida paralelne sopstvenim vektorima kovarijansne matrice ulaznog signala kada je srednja vrednost ulaznog signala nula.

U slučaju kada se primenjuje PSA/MSA algoritam, sinaptički vektori, koji su na slici prikazani isprekidanim linijama, se ne poklapaju sa osama hiperelipsoida. Oni su za izvestan ugao rotirani u odnosu na njih. Vektori su međusobno ortonormalni pod uticajem “sile” PSA algoritma – ta “sila” je na slici označena sa F_{PSA}. Sada se postavlja pitanje koju dodatnu “silu” treba primeniti da bi se sinaptički vektori dodatno rotirali do osa hiperelipsoida. Ovde će biti predloženo da se u PSA/MSA algoritam doda još jedan član. Taj novi član bi imao zadatak da doda “diferencijalni momenat” koji je na slici označen “silama” F_{1ILA} i F_{2ILA}.



Slika 1 Gustina verovatnoće multiltivarijabilnog Gausovog signala

Za realizaciju “porodičnog principa” predlažemo sledeći opšti metod, koji transformiše PSA/MSA algoritam, označen sa $FLA_{PSA/MSA}$ u PCA/MCA algoritam, označen sa $LA_{PCA/MCA}$:

$$\mathrm{LA}_{\mathrm{PCA/MCA}} = \mathrm{FLA}_{\mathrm{PSA/MSA}} + \alpha\, \mathrm{ILA}\left(\max\left(\mathrm{E}\left\{ \left(y^{\mathrm{T}} y \right) \middle| \begin{array}{l} w_k^{\mathrm{T}} w_k = 1, \\ k = 1,2\ldots N \end{array} \right\} \right) \right), \tag{5}$$

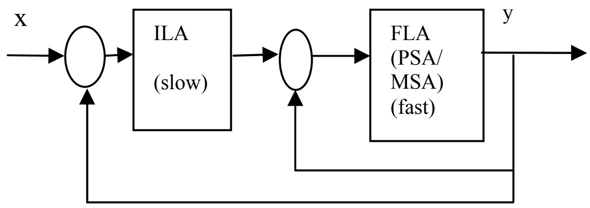
gde je α konstanta takva da važi $|\alpha|<1$. ILA označava individualni deo zakona učenja. ILA predstavlja algoritam za maksimizaciju $E(y^T y)$ pod ograničenjem $w_k^T w_k=1$ za $k=1,2,\ldots,N$. Drugim rečima, novi PCA/MCA algoritam učenja ima sledeću formu

$$\Delta W_{\mathrm{PCA/MSA}} = \Delta W_{\mathrm{FLA}} + \alpha \Delta W_{\mathrm{ILA}}, \tag{6}$$

gde ΔW_{FLA} označava doprinos porodičnog dela zakona učenja, dok ΔW_{ILA} označava doprinos individualnog dela zakona učenja. Lako je uočiti da ako se koristi homogeni PSA/MSA algoritam rezultujući PCA/MCA algoritam će takođe biti homogen. Ovaj metod ćemo nazvati vremenski orijentisan hijerarhijski metod (TOHM).

Pošto je $|\alpha| < 1$, porodični deo zakona učenja se primenjuje brže nego individualni deo zakona učenja. Jednačina (6) je uprošćeno prikazana na slici 2.

Iako je usvojeno $|\alpha| < 1$, sada ćemo se pozabaviti uprošćenom analizom u slučaju kada je $|\alpha|$ mnogo manje od 1. U tom slučaju, individualni deo jednačine (6) skoro da nema nikakvog uticaja na porodični deo. Tada porodični deo obezbeđuje dostizanje



Slika 2 Blok šema realizacije predloženog metoda

glavnog/sporednog potprostora. To znači da je $W^T W$=I i da W pokriva glavni/sporedni potprostor. Drugim rečima, usled uvođenja dve vremenske skale možemo pristupiti problemu PCA/MCA kao optimizacionom problemu. U suštini, onaj deo zakona učenja koji se odvija na bržoj vremenskoj skali, i koji obavlja PSA/MSA, predstavlja ograničenje za sporiji (individualni) deo zakona učenja. Znači, ako usvojimo da je PSA/MSA deo zakona učenja dovoljno brz može se reći da on predstavlja striktno ortonormalno ograničenje na PSA/MSA potprostoru. U tom slučaju, za individuani deo u jednačini (6) možemo pisati

$$\begin{aligned}&W(i+1) = W(i) + \gamma(i)\left(x(i)y(i)^{\mathrm{T}} - W(i)diag\left(y_1(i)^2, \cdots, y_N(i)^2\right)\right)\alpha,\\&\text{gde je } \alpha \text{ takvo da je } |\alpha| < 1,\\&W^{\mathrm{T}}W = \mathrm{I} \text{ i pokriva glavni/sporedni podprostor.}\end{aligned} \tag{7}$$

Odgovarajuća diferencijalna jednačina je [17, 18]

$$\frac{dW}{dt} = \left(CW - Wdiag\left(W^{\mathrm{T}}CW\right)\right)\alpha\,, \tag{8}$$

gde diag($W^{\mathrm{T}}CW$) predstavlja dijagonalnu matricu koja se dobija kada se svi elementi van glavne dijagonale matrice $W^{\mathrm{T}}CW$ anuliraju. Ako napišemo jednačinu za pojedinačne kolone w_k, imamo

$$\frac{dw_k}{dt} = \alpha\left(Cw_k - \lambda_k w_k\right), \tag{9}$$

gde je λ_k k-ti dijagonalni element matrice diag ($W^{\mathrm{T}}CW$). Možemo lako zaključiti da su stacionarne tačke ove jednačine sopstveni vektori matrice C. Pošto w_k pokrivaju dominantni/sporedni potprostor matrice C, onda su jedino moguća rešenja dominantne/sporedne komponente. Ako $w_k(i)$ pridružene diferencne jednačine posećuje beskonačno često skoro sigurno, kompaktan podskup domena privlačenja rešenja jednačine (9), tada rešenje jednačine (9) predstavlja rešenje diskretne jednačine (7). Naravno, veoma je teško (ako ne i nemoguće) dokazati ovo u opštem slučaju.

Usled simetrije predloženog algoritma, sistem ima više od jednog rešenja. To može prouzrokovati "konflikt interesa" između pojedinih neurona. Ako razmotrimo slučaj na slici 1 može se uočiti da dok se jedan sinaptički vektor kreće u pravcu pozitivnog dela dominantne ose 1, u isto vreme drugi sinaptički vektor može da ima "nameru" da se kreće ka negativnom delu dominantne ose 1. To znači da algoritam može da postane osetljiv na izbor koeficijenta α. Osim toga, značajno je proučiti i slučaj u kome realizacija algoritma u paralelnom hardveru dovodi do toga da koeficijent α nije savršeno jednak za sve vektore, pa se u tom slučaju postavlja pitanje stabilnosti algoritma.

Sada ćemo analizirati slučaj kada je u algoritam uneta nekakva vrsta asimetrije. Drugim rečima, daćemo pojedinim vektorima drugačije šanse za dostizanje dominantnog/najsporednijeg sopstvenog vektora. U tom slučaju, predloženi metod se može opisati sledećom jednačinom:

$$\mathrm{LA}_{\mathrm{PCA/MCA}} = \mathrm{FLA}_{\mathrm{PSA/MSA}} + \mathrm{ILA}\left(\max\left(\mathrm{E}\left\{\left((Dy)^{\mathrm{T}} y\right) \middle| \begin{array}{l} w_k^{\mathrm{T}} w_k = 1, \\ k = 1,2\ldots N \end{array}\right\}\right)\right), \tag{10}$$

gde je D dijagonalna matrica sa nenultim dijagonalnim elementima d_n takvim da je $|d_n|<1$. Unošenje asimetrije dovodi do dodatne vremenske hijerarhije u realizaciji individualnih delova zakona učenja. Ako su svi d_n jednaki α, ova jednačina se svodi na homogeni slučaj.

Ponovo ćemo sprovesti uprošćenu analizu za slučaj kada je $|d_n|$ mnogo manje od 1. U tom slučaju, individualni deo u jednačini (10) praktično nema uticaj na porodični deo. Tada porodični deo obezbeđuje pronalaženje glavnog/sporednog potprostora. To znači da je $W^T W$=I i da W spans glavni/sporedni potprostor. Kao što je već ranije objašnjeno možemo pristupiti problemu PCA/MCA kao optimizacionom problemu. U stvari brži deo zakona učenja, koji obezbeđuje izračunavanja PSA/MSA predstavlja ograničenje za sporiji (individualni) deo zakona učenja. To znači da ako pretpostavimo da je PSA/MSA deo zakona učenja dovoljno brz možemo reći da u tom slučaju taj deo zakona učenja predstavlja ortonormalno ograničenje na PSA/MSA potprostoru. U tom slučaju, individualni deo u jednačini (10) možemo pisati kao

$$\begin{array}{l} W(i+1) = W(i) \\ + \gamma(i)\left(x(i)y(i)^{\mathrm{T}} - W(i)diag\left(y_1(i)^2, \cdots, y_N(i)^2\right)\right)D, \\ \text{gde je } D \text{ dijagonalna matrica } sa \text{ dijagonalnim elementima } d_k \\ \text{takvim da je } |d_k| < 1 \text{ i} \\ W^{\mathrm{T}} W = \mathrm{I} \text{ i pokriva glavni/sporedni podprostor.} \end{array} \tag{11}$$

Odgovarajuća diferencijalna jednačina je [17, 18]

$$\frac{dW}{dt} = \left(CW - Wdiag\left(W^{\mathrm{T}} CW\right)\right)D, \tag{12}$$

gde je diag($W^T CW$) definisana kao i ranije. Ako napišemo jednačinu za svaku kolonu w_k, imamo

$$\frac{dw_k}{dt} = d_k \left(Cw_k - \lambda_k w_k\right), \tag{13}$$

gde je λ_k k-ti dijagonalni element matrice diag($W^{\mathrm{T}}CW$). Nije teško zaključiti da su stacionarne tačke ove jednačine sopstveni vektori matrice C. Ako $w_k(i)$ pridruženog diskretnog algoritma posećuje beskonačno često skoro sigurno kompaktni podskup domena privlačenja jednačine (13), tada rešenje jednačine (13) predstavlja i rešenje algoritma (10). Ponovo ćemo naglasiti, da je veoma teško pokazati tu osobinu diskretnog algoritma u opštem slučaju.

3. MATEMATIČKA ANALIZA PREDLOŽENOG PRINCIPA

U ovom paragrafu će biti pokazano kako predloženi TOHM metod može biti predstavljen i kao učenje na približno Stifel-Grasmanovoj (Stiefel-Grassman) podmnogostrukosti. Prvo će biti predstavljen algoritam koji izvodi učenje na Stifel-Grasmanovoj podmnogostrukosti i kasnije će biti pokazano kako se on može povezati sa TOHM metodom. Ovde ćemo se uglavnom baviti algoritmima koji predstavljaju zakone učenja na Grasmanovoj podmnogostrukosti dok se sve definicije mogu izvesti analogno za zakone učenja na Stifelovoj podmnogostrukosti.

Prvo ćemo neformalno ponoviti definiciju Grasmanove mnogostrukosti koja je data u [19], [20]:

Prostor matrica $W \in O^{KxN} \subset R^{KxN}$ $(N \leq K)$ takvih da je $W^T W$=I i homogena funkcija $J: O^{KxN} \rightarrow$ R takva da je $J(W)=J(WQ)$ za bilo koju NxN ortogonalnu matricu Q se nazivaju Grasmanovom mnogostrukošću.

Algoritmi učenja koji se primenjuju u neuralnim mrežama se često mogu shvatiti kao algoritmi koji maksimiziraju/minimiziraju funkciju cene pod nekakvim ograničenjem, pri čemu je to ograničenje najčešće ortonormalnost sinaptičke matrice.

Takvi zakoni učenja predstavljaju rešenje sledećeg problema:

Nađi W_{eks} takvu da važi: $J(W_{eks})$ = max/min $J(W)$, $W \in R^{KxN.}$

Standardan način za dobijanje željenog rešenja je definisanje Lagranžove funkcije:

$$J_l(W) = J(W) + l\,\mathrm{tr}\left(W^{\mathrm{T}} W - \mathrm{I}\right)$$

gde je l takozvani Lagranžov multiplikator, i traženje ekstremuma funkcije $J_l(W)$, na primer upotrebom tehnike uzlaznih/silaznih gradijenata,

$$\frac{dW}{dt} = \eta \nabla J_l .$$

Kako bi se osigurala ortonormalnost sinaptičkih kolona, može se primeniti iterativna ortogonalizacija kolona matrice W, na primer upotrebom Gram-Schmidt metoda ortogonalizacije matrice čiji se elementi menjaju na osnovu gradijentne optimizacije funkcije J(W) ili pomoću projekcije na ortonormalnu grupu [21]. Međutim, iterativna primena ortonormalnosti može biti problematična u praksi ([21]). To je razlog zašto su istraživači počeli da proučavaju zakone učenja koji održavaju ortonormalnost sinaptičke matrice u svakom momentu. Takvi algoritmi su poznati kao SOC (Orthonormal Strongly-Constrained – striktno ortnormalno ograničeni) algoritmi. Nekoliko algoritama iz te klase je analizirano u [21].

Prvo analizirajmo sledeći algoritam:

$$\mathrm{E}(y^2) = \max \tag{14}$$

$$y = \frac{w^{\mathrm{T}}}{\sqrt{w^T w}} x \tag{15}$$

gde x predstavlja ulazni vektor za jednoslojnu mrežu sa jednim izlaznim neuronom (y) i w predstavlja sinaptički vektor. Jednostavno je uočiti da jednačinu (14) možemo napisati i kao

$$\mathrm{E}(y^2) = \frac{w^{\mathrm{T}} x x^{\mathrm{T}} w}{w^{\mathrm{T}} w} = \max \tag{16}$$

Ako sada direktno generišemo uzlazni gradijentni algoritam u odnosu na y^2 pri čemu se gradijent računa po w, imamo:

$$w(i+1) = w(i) + \gamma(i) \frac{xy\sqrt{w(i)^{\mathrm{T}} w(i)} - y^2 w(i)}{w(i)^{\mathrm{T}} w(i)}. \tag{17}$$

Jednostavno se uočava da zamenom $w(i)^{\mathrm{T}} w(i)=1$ dobijamo direktno poznati Ojin zakon učenja [5]. Ako generalizujemo ovaj algoritma na slučaj sa više izlaznih neurona

$$\mathrm{E}(\mathrm{tr}(yy^{\mathrm{T}})) = \max$$

$$y_k = \frac{w_k^{\mathrm{T}}}{\sqrt{w_k^{\mathrm{T}} w_k}} x,$$

gde y_k predstavlja k-ti izlazni neuron i w_k predstvalja k-tu kolonu sinaptičke matrice W, zakon učenja za modifikaciju sinaptičkih vektora je u tom slučaju

$$w_k(i+1) = w_k(i) + \gamma(i) \frac{xy_k\sqrt{w_k(i)^{\mathrm{T}} w_k(i)} - y_k^2 w_k(i)}{w_k(i)^T w_k(i)}. \tag{18}$$

Sada ćemo definisati Grasmananovu dominantnu podmnogostrukost.

Prostor matrica $W \in O^{KxN} \subset R^{KxN}$ $(N \leq K)$ takvih da je $W^T W=\mathrm{I}$ i W pokriva dominantni potprostor matrice C, i funkcija $J: O^{KxN} \to R$ takva da je $J(W)=J(WQ)$ za bilo koju NxN ortonormalnu matricu Q ćemo nazvati Grasmanovom dominantnom podmnogostrukošću.

Ako sada primenimo algoritam (18) pod striktnim ograničenjem za $W - W$ je takvo da je $W^T W=\mathrm{I}$ i W pokriva dominantni potprostor matrice C,u stvari imamo sledeći algoritam

$$w_k(i+1) = w_k(i) + \gamma(i)\left(xy_k - y_k^2 w_k(i)\right),$$

$$y_k = w_k^{\mathrm{T}} x.$$

Funkcija cene u kompaktnoj notaciji je data sa

$$E(\mathrm{tr}(yy^{\mathrm{T}})) = E(\mathrm{tr}(W^{\mathrm{T}}xx^{\mathrm{T}}W) = \mathrm{tr}(W^{\mathrm{T}}CW) = \max .$$

Nije teško uočiti da se radi o funkciji za koju je ispunjeno *J(WQ)=J(W)*. Drugim rečima možemo reći algoritam (18) vrši učenje na Grasmanovoj dominantnoj podmnogostrukosti. Algoritmu (18) se može pridružiti sledeća diferencijalna jednačina (u kompaktnoj notaciji)

$$\frac{dW}{dt} = \left(CW - Wdiag\left(W^{\mathrm{T}}CW\right)\right), \tag{19}$$

gde je diag$(W^{\mathrm{T}}CW)$ dijagonalna matrica definisana kao i ranije. Ako napišemo (19) za svaku kolonu w_k, imamo

$$\frac{dw_k}{dt} = \left(Cw_k - \lambda_k w_k\right), \tag{20}$$

gde je λ_k k-ti element glavne dijagonale matrice diag$(W^{\mathrm{T}}CW)$. Lako je zaključiti da su stacionarne tačke ovih jednačina sopstveni vektori matrice C. Pošto se učenje obavlja na dominantnoj Grasmanovoj podmnogostrukosti rezultujući vektor w_k će biti jednak nekom od dominantnih sopstvenih vektora matrice C. Ako $w_k(i)$ pridruženog diskretnog algoritma posećuje beskonačno često kompaktan podskup domena privlačenja rešenja jednačine (20), tada su rešenja jednačine (20) takođe rešenja pridruženog diskretnog algoritma (15). Naravno, kao što je već rečeno, ovu osobinu algoritma (15) je teško dokazati.
Na sličan način se može definisati Grasmanova sporedna podmnogostrukost.

Prostor matrica $W \in O^{KxN} \subset R^{KxN}$ $(N \leq K)$ takvih da je $W^{\mathrm{T}}W$=I i W pokriva sporedni potprostor matrice C, i funkcija $J\colon O^{KxN} \rightarrow R$ takva da je J(W)=J(WQ) za bilo koju NxN ortonormalnu matricu Q će bit nazvani Grasmanovom sporednom podmnogostrukošću.

Ako se algorithm (18) primeni pod striktnim ograničenjem za $W - W$ je takvo da je $W^{\mathrm{T}}W$=I i W pokriva sporedni potprostor matrice C, i primenimo analizu kao u slučaju dominantne Grasmanove podmnogostrukosti kolone sinaptičke matrice W će biti jednake sporednim sopstvenim vektorima matrice C.

U praksi je problem što se u toku rada algoritma ne može tačno odrediti Grasmanova podmnogostrukost. Zbog toga se mora pribeći primeni Lagranžovog metoda, čija primena rezultuje sledećim zakonom učenja

$$w_k(i+1) = w_k(i) + \gamma(i)\left(xy_k - y_k{}^2 w_k(i)\right) + \beta\gamma(i)\Delta W_{\mathrm{PSA/MSA}}, \tag{21}$$

gde $\Delta W_{PSA/MSA}$ reprezentuje deo zakona učenja koji je dat usvojenim PSA/MSA zakonom učenja i koji reprezentuje “slabo” ograničenje. U tom slučaju jednačina (10) se može pisati u sledećem obliku

$$w_k(i+1) = w_k(i) + \alpha\gamma(i)\left(xy_k - y_k{}^2 w_k(i)\right) + \gamma(i)\Delta W_{\mathrm{PSA/MSA}}. \tag{22}$$

Ako je α dovoljno malo možemo smatrati da deo zakona učenja koji se množi sa α ne utiče na PSA/MSA deo zakona učenja tako da imamo približno učenje na Grasmanovom glavnom/sporednom potprostoru. Ako je α dovoljno malo, možemo smatrati da je analiza koja je izvedena za SOC algoritme validna i za algoritam (22). Ovde se mora naglasiti da se analiza stabilnosti algoritma mora sprovesti posebno za svaki partikularni izbor PSA/MSA algoritma. Tek posle takve analize je moguće odrediti odgovarajuću vrednost parametra α.

4. ZAKLJUČAK

U radu je prikazan opšti metod koji vrši transformaciju PSA/MSA metoda u PCA/MCA metode. Algoritam je baziran na vremenski orijentisanom hijerarhijskom principu. Uvođenje dve vremenske skale je glavna novost predloženog metoda. To indirektno znači da moguća biološka implementacija neuralne mreže zahteva postojanje dva tipa neurotransmitera. Na bržoj vremenskoj skali, PSA/MSA algoritam je odogovoran za "ponašanje" svih izlaznih neurona (porodicu). Na sporijoj vremenskoj skali, izlazni neuroni se takmiče za "ispunjenje sopstvenih interesa". Na toj vremenskoj skali se vektori usmeravaju ka glavnim komponentama ulazne kovarijansne matrice. U radu je prikazana pojednostavljena matematička analiza u opštem slučaju.

LITERATURA

[1] Pearson, K.. "On lines and planes of closest fit to systems of points in space". *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, Sixth Series 2, 1901, 559--572.

[2] Hotelling, H. "Analysis of a complex of statistical variables into principal components". *Journal of Educational Psychology*, Vol. 24, 1933, pp. 417-441,498-520.

[3] Loéve, M. "Fonctions al´eatoires du second ordre," in Processus stochastiques et mouvements Browniens (P. L´evy, ed.), 1948, Paris, France: Gauthier-Villars.

[4] Amari, S-I. "Neural theory of association and concept formation", *Biological Cybernetics*, Vol. 26, 1977, pp. 175-185.

[5] Oja, E. "A Simplified neuron model as a principal component analyzer", *J. Math. Biol.*, vol. 15, 1982, pp. 267-273.

[6] Oja, E., Karhunen, J. and Hyvarinen, A. "From neural principal components to neural independent components", *Int. Conference on Arrtifical Neural Networks,* 1997, Lausanne, Switzerland.

[7] Bannour, S. and Azimi-Sadjadi, M.R. "Principal component extraction using recursive least squares learning", *IEEE Trans. On Neural Networks*, Vol. 6, 1989, pp.456-469

[8] Cichocki, A., Gharieb, R.R. and Hoya, T. "Efficient extraction of evoked potentials by combination of independent component analysis and cumulant-based matched filtering", *11th IEEE Signal Processing Workshop on Statistical Signal Processing*, 2001, pp. 237-240, Singapore.

[9] Diamantaras, K.I. and Kung, S.Y., *Principal Component Neural Networks – Theory and Applications* 1996, John Wiley & Sons Inc., New York.

[10] Turk, M. and Pentland, A. "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, Vol. 3, 1991, pp. 71-86.

[11] Kung, S.Y., Diamantaras, K. and Taur, J. “Neural networks for extracting pure/constrained/oriented principal components”, in J.R. Vaccaro, editor, SVD and Signal Processing, 1991, pp. 57-81, Elsevier Science, Amsterdam.

[12] Klemm, R. “Adaptive airborne MTI: an auxiliary channel approach”, *IEE Proceedings*, Vol. 134, 1987, pp. 269-276.

[13] Mathew, G. and Reddy, V. “Orthogonal eigensubspace estimation using neural networks”, *IEEE Trans. On Signal Processing*, Vol. 42, 1994, pp. 1803-1811.

[14] Schmidt, R. “Multiple emitter location and signal parameter estimation”, *IEEE Trans. On Antennas and Propagation*, Vol. 34, 1986, pp. 276-280.

[15] Wiscott, L., “Learning invariance manifolds”, *International Conference on Artificial Neural Networks (ICANN)*, 1998, pp. 555-560, 1998.

[16] Xu, L., Oja, E. and Suen, C.Y. “Modified Hebbian learning for curve and surface fitting”, *Neural Networks,* Vol. 5, 1992, pp. 441-457.

[17] Ljung, L.“Analysis of recursive stochastic algorithms”, *IEEE Trans. Automat. Contr*., vol. 22, 1977, pp. 551-575.

[18] Oja, E. and Karhunen, J. “On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix”, *J. Math. Anal., Appl*., 106, 1985, pp. 69-84.

[19] Edelman, A., Arias, T.A., and Smith, S.T.“The geometry of algorithms with orthogonality constraints”, *SIAM Journal on Matrix Analysis Applications,* Vol. 20, 1998, No. 2, pp. 303-353.

[20] Fiori, S. ”A minor subspace algorithm based on neural Stiefel dynamics”, *International Journal of Neural Systems*, Vol. 12, No. 5, 2002, pp. 339 - 350.

[21] Fiori, S. "A theory for learning by weight flow on Stiefel-Grassman Manifold”, *Neural Computation,* Vol. 13, 2001, pp. 1625-1647.

Abstract: This paper presents first general method which transforms neural networks that are used for calculation of principal/minor subspace (PSA/MSA) into neural networks for calculation of principal/minor components (PCA/MCA). By using this method it is possible to create neural networks that are suitable for realization in parallel hardware since it is possible to construct networks (algorithms) that are homogeneuous from the point of view of individual neuron.

Key words: *principal component analysis, minor component analysis, principal subspace analysis, minor subspace analysis, time oriented hierarchical method, neural networks*

TIME ORIENTED HIERARCHICAL METHOD FOR CALCULATION OF PCA AND MCA

Marko Janković