

Dekompozicije tenzora i primjena u izdvajanju značajki

Ante Jukić

Zavod za laserska i atomska istraživanja i razvoj, Institut Ruđer Bošković
ajukic@irb.hr

Sažetak – Dekompozicije tenzora su važan alat u modeliranju i analizi višedimenzionalnih podataka i signala. Posebno su važne dvije vrste modela koje možemo smatrati poopćenjem dekompozicije singularnih vrijednosti: kanonski poliadski model (CPD) i Tucker model. U ovom radu su opisani navedeni modeli i odgovarajući algoritmi za dekompoziciju. Prikazana je i dekompozicije tenzora na problemima izdvajanja značajki za klasifikaciju bolesti iz proteinskih profila i klasifikaciji melanoma iz RGB slika.

Ključne riječi – Dekompozicija tenzora, multilinearna analiza, analiza višedimenzionalnih podataka.

I. UVOD

U modernim primjenama često se pojavljuju višedimenzionalni podaci. Prirodan prikaz takvih podataka su tenzori, koje možemo smatrati višedimenzionalnim poljima. Pojam tenzora je poopćenje pojma vektora i matrice, pa je time i analiza podataka temeljena na tenzorima poopćenje dvodimenzionalnih pristupa, odnosno metoda za faktorizaciju matrica kao što su faktorska analiza, analiza nezavisnih komponenti, faktorizacija nenegativnih matrica ili analiza rijetkih komponenti. Podaci prikupljeni raznim eksperimentima prirodno tvore višedimenzionalna polja podataka, primjerice u psihometriji, neuroznanosti, kemometriji, računalnom vidu ili analizi društvenih mreža. Takvih podatci se često dekomponiraju u odgovarajući tenzorski model čime se dobiva bolja interpretacija latentnih faktora. Pristup opisa podataka modelom manje kompleksnosti je čest u rješavanju raznih problema, poput uklanjanja šuma, smanjenja dimenzionalnosti, slijepe separacije signala, strojnog učenja ili raspoznavanja uzoraka [1].

Dekompozicije tenzora vuku korijene u radovima Hitchcocka iz 1927. godine [2], [3], dok je Catell donio prve ideje za modele višeg reda 1944. godine [4], [5]. Navedeni radovi nisu bili primijećeni sve do 60ih i 70ih godina kada su Tucker [6], [7], [8], Harshman [9] i Carroll i Chang [10] objavili važne radove u području analize tenzora. Svi navedeni radovi objavljeni su u psihometrijskim časopisima, ali im je vidljivost rasla i tenzori su se počeli koristiti i u drugim područjima kao što je primjerice kemometrija [11], [12]. Posljednjih godina tenzori su postali iznimno popularni u analizi signala i

slika, numeričkoj linearnoj algebri, računalnom vidu, numeričkoj analizi, analizi grafova, komunikacijama i drugim područjima. Posebno zanimljivo područje primjene zadnjih godina je u odabiru značajki. Dekompozicije tenzora se koriste u predobradi podataka prije klasifikacije kao metoda za izdvajanje diskriminativnih latentnih značajki i istovremeno smanjenje dimenzije uzoraka [13], [14], [15], [16], [17], [18], [19].

Modeli koji se najčešće koriste su kanonska poliadaska dekompozicija (eng. *Canonical Polyadic Decomposition*, CPD) i Tuckerova dekompozicija koji se mogu promatrati kao poopćenje dekompozicije singularnih vrijednosti (eng. *Singular Value Decomposition*, SVD) i analize glavnih komponenti (eng. *Principal Component Analysis*, PCA). U praksi se primjenjuju u raznim varijantama uz dodatna ograničenja kao što su primjerice nenegativnost ili rijetkost pa ne vrijedi potpuna analogija sa SVD ili PCA, ali se pokazuju iznimno korisnima [1]. Također postoje razni programski paketi za rad sa tenzorima [15], [20], [21], [22], [23]. Detaljan pregled dekompozicija tenzora i njihovih primjena može se naći u [24], [1], [25].

U ostatku rada dan je pregled osnovnih pojmova vezanih za tenzore i elementi tenzorske algebre. Predstavljani su osnovni modeli za dekompozicije tenzora i njihove varijacije te su opisani tipični algoritmi koji se koriste u analizi. Dani su primjeri primjene dekompozicija tenzora u izdvajanju značajki.

II. TEORIJA

U ovoj sekciji je dan kratki pregled osnovnih pojmova multilinearne algebre, te su opisani modeli tenzora i tipični algoritmi za dekompoziciju. U nastavku ćemo skalare označavati malim slovima (npr. x), vektore malim podebljanim slovima (npr. $\mathbf{x}$), matrice velikim podebljanim slovima (npr. $\mathbf{X}$), a tenzore višeg reda velikim podebljanim i podcrtanim slovima (npr. $\underline{\mathbf{X}}$).

A. Osnovi pojmovi i računanje s tenzorima

Tenzor je višedimenzionalni ekvivalent vektora i matrice i možemo ga predstaviti višedimenzionalnim poljem brojeva. Red tenzora jednak je broju indeksa potrebnih za indeksiranje njegovih elemenata. Primjerice, skalar možemo smatrati tenzorom reda 0, vektor tenzorom reda 1 a matricu tenzorom reda 2. Tenzori reda većeg ili jednakog 3 nazivaju se tenzorima višeg reda. Formalno,

tenzor reda N možemo definirati kao N -linearnu formu. Neka su $\mathbb{V}^{(1)}, \dots, \mathbb{V}^{(N)}$ konačnodimenzionalni vektorski prostori nad poljem $\mathbb{K}$, sa dimenzijama $I_1, \dots, I_N$. Tada elemente vektorskog prostora $\mathbb{V}^{(1)} \circ \dots \circ \mathbb{V}^{(N)} = \circ_{n=1}^N \mathbb{V}^{(n)}$ nazivamo tenzorima, gdje $\circ$ označava tenzorski (vanjski) produkt vektorskih prostora¹. Elemente tog vektorskog prostora dimenzije $\prod_{n=1}^N I_n$ nazivamo tenzorima. Tada se svaki tenzor $\mathcal{T}$ može se prikazati u obliku

$$\mathcal{T} = \sum_{i_1=1}^{I_1} \dots \sum_{i_N=1}^{I_N} t_{i_1 \dots i_N} \mathbf{e}_{i_1}^{(1)} \circ \dots \circ \mathbf{e}_{i_N}^{(N)},$$

gdje su $\{\mathbf{e}_i^{(n)}\}$ baze svakog od vektorskih prostora $\mathbb{V}^{(n)}$, a $t_{i_1 \dots i_N} \in \mathbb{K}$ koordinate tenzora $\mathcal{T}$ u odabranim bazama. Uz fiksni odabir baza svaki tenzor možemo jednoznačno predstaviti N -dimenzionalnim poljem elemenata iz $\mathbb{K}$

$$\mathbf{T} = (t_{i_1 \dots i_N}) \in \mathbb{K}^{I_1 \times \dots \times I_N} \quad (1)$$

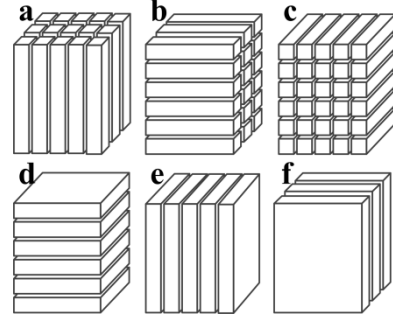
čije su komponente koordinate tenzora. Ovaj način predstavljanja tenzora ponekad se naziva hipermatricom [26]. Ako napravimo promjenu baze na prostorima $\mathbb{V}^{(n)}$, definiranu sa $\mathbf{e}_i^{(n)} = \mathbf{Q}^{(n)} \tilde{\mathbf{e}}_i^{(n)}$ za neke matrice $\mathbf{Q}^{(n)}$, $n \in \{1, \dots, N\}$ tada dobivamo nove koordinate tenzora $\mathcal{T}$ koje možemo zapisati preko koordinata u početnoj bazi kao

$$\tilde{t}_{j_1 \dots j_N} = \sum_{i_1=1}^{I_1} \dots \sum_{i_N=1}^{I_N} Q_{j_1 i_1}^{(1)} \cdot \dots \cdot Q_{j_N i_N}^{(N)} \cdot t_{i_1 \dots i_N},$$

pa vidimo da drugačiji odabir baza znači i drugi prikaz tenzora² [27]. Iz navedenoga je jasno da tenzor prirodno definira multilinearni operator a njegov zapis (1) opisuje djelovanje tog operatora u određenoj N -torci baza, na isti način kao što matrica definira linearni operator između dva vektorska prostora u zadanom paru baza [28]. Međutim, u ovom radu tenzori se promatraju u prvom redu kao višedimenzionalna struktura podataka. Stoga ćemo u nastavku pod pojmom tenzor podrazumijevati strukturu u obliku (1), i promatrati ćemo isključivo realne tenzore, tj. tenzore čije su komponente realni brojevi.

Svaki indeks u tenzoru nazivamo mod, a dimenzija pripadnog moda je broj različitih vrijednosti koje taj indeks može poprimiti [29]. Unutar tenzora možemo indeksirati podtenzore ograničavanjem pojedinih indeksa. Primjerice, za tenzor $\mathbf{X} = (x_{i_1 i_2 i_3}) \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ reda 3 možemo fiksiranjem indeksa u modu 1, 2 ili 3 definirati podtenzore kao $\mathbf{X}_{i_1=n} = \mathbf{X}_{n::}$, $\mathbf{X}_{i_2=n} = \mathbf{X}_{:n:}$ ili $\mathbf{X}_{i_3=n} = \mathbf{X}_{::n}$. Primjetimo da su u tom slučaju dva indeksa slobodna pa su podtenzori zapravo novi tenzori reda 2, odnosno matrice. Posebno za slučaj tenzora reda 3 te podtenzore nazivamo horizontalnim, lateralnim i frontalnim odsječcima. Također možemo definirati i vektore po modu n na način da fiksiramo sve indekse osim jednoga, npr. vektor u modu 1 se dobije kao $\mathbf{x}_{:i_2 i_3}$. Analogno za tenzor reda N možemo

definirati podtenzore reda manjeg ili jednakog N . Na slici Slika 1 su prikazani primjeri podtenzora za tenzor reda 3.



Slika 1: Primjeri podtenzora. Vektori tenzora u (a) modu 1, (b) modu 2, (c) modu 3; i (d) horizontalni, (e) lateralni i (f) frontalni odsječci.

Često je korisno prikazati tenzor u obliku matrice. Stoga definiramo matricizaciju tenzora po modu n kao matricu $\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times \prod_{k=1, k \neq n}^N I_k}$ koja se dobije tako da vektore iz moda n složimo kao stupce u matricu $\mathbf{X}_{(n)}$. Poredak kojim se vektore iz moda n preslikavaju u stupce nije bitan, dok god se isti poredak koristi u svim izračunima [25], [1]. Na primjer, matricizacijom tenzora trećeg reda po modu 3 dobivamo matricu $\mathbf{X}_{(3)} \in \mathbb{R}^{I_3 \times I_1 \cdot I_2}$ sa stupcima jednakima $\mathbf{x}_{i_1 i_2 :}$. Ako matricu $\mathbf{A} \in \mathbb{R}^{I \times J}$ promatramo kao tenzor drugog reda tada vrijedi $\mathbf{A}_{(1)} = \mathbf{A}$, $\mathbf{A}_{(2)} = \mathbf{A}^T$.

Skalarni produkt dva tenzora istog reda i jednakih dimenzija $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ definira se kao

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} X_{i_1 i_2 \dots i_N} \cdot Y_{i_1 i_2 \dots i_N},$$

dok je norma tenzora inducirana tim skalarnim produktom jednaka $\|\mathbf{X}\| = \sqrt{\langle \mathbf{X}, \mathbf{X} \rangle}$. Ovako definirana norma kompatibilna sa Frobeniusovom normom u matričnom slučaju. Također je korisno primijetiti da vrijedi $\|\mathbf{X}\| = \|\mathbf{X}_{(n)}\|_F$ za bilo koji $n \in \{1, \dots, N\}$, gdje je $\|\cdot\|_F$ Frobeniusova norma.

U radu s tenzorima pojavljuju se razne vrste produkata tenzora, matrica i vektora. Vanjskim produktom tenzora $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ i $\mathbf{Y} \in \mathbb{R}^{J_1 \times J_2 \times \dots \times J_M}$ dobivamo novi tenzor

$$\mathbf{Z} = \mathbf{X} \circ \mathbf{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N \times J_1 \times J_2 \times \dots \times J_M}$$

čiji su elementi dani sa

$$Z_{i_1 i_2 \dots i_N j_1 j_2 \dots j_M} = x_{i_1 i_2 \dots i_N} \cdot y_{j_1 j_2 \dots j_M}.$$

Poseban slučaj je vanjski produkt dva vektora $\mathbf{a} \in \mathbb{R}^I$ i $\mathbf{b} \in \mathbb{R}^J$ čime dobivamo matricu ranga 1

$$\mathbf{A} = \mathbf{a} \circ \mathbf{b} = \mathbf{a} \mathbf{b}^T \in \mathbb{R}^{I \times J}.$$

Općenito, vanjskim produkt N vektora $\mathbf{a}^{(n)} \in \mathbb{R}^{I_n}$, $n \in \{1, \dots, N\}$ dobivamo tenzor

¹ Vektorski prostor $\mathbb{V}^{(1)} \circ \dots \circ \mathbb{V}^{(N)}$ možemo zamisliti kao skup svih linearnih kombinacija tenzorskih produkata elemenata iz svakog od prostora.

² Ovaj multilinearni izraz se u literaturi često zapisuje kao $\tilde{\mathbf{T}} = (\mathbf{Q}^{(1)}, \dots, \mathbf{Q}^{(N)}) \cdot \mathbf{T}$.

$$\underline{\mathbf{Z}} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \dots \circ \mathbf{a}^{(N)} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N} \quad (2)$$

sa elementima $z_{i_1 i_2 \dots i_N} = a_{i_1}^{(1)} \cdot a_{i_2}^{(2)} \cdot \dots \cdot a_{i_N}^{(N)}$. Za tenzor koji možemo prikazati kao vanjski produkt vektora kažemo da je tenzor ranga jedan [25], [1]. Kroneckerov produkt matrica $\mathbf{A} \in \mathbb{R}^{I \times J}$ i $\mathbf{B} \in \mathbb{R}^{T \times R}$ je matrica

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \dots & a_{1J}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{I1}\mathbf{B} & \dots & a_{IJ}\mathbf{B} \end{bmatrix} \in \mathbb{R}^{IT \times JR}.$$

Khatri-Rao produkt se definira za matrice koje imaju jednak broj stupaca kao

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}^{(1)} \otimes \mathbf{b}^{(1)}, \mathbf{a}^{(2)} \otimes \mathbf{b}^{(2)}, \dots, \mathbf{a}^{(J)} \otimes \mathbf{b}^{(J)}] \in \mathbb{R}^{IT \times J}$$

gdje su $\mathbf{A} = [\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \dots, \mathbf{a}^{(J)}] \in \mathbb{R}^{I \times J}$ i $\mathbf{B} = [\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \dots, \mathbf{b}^{(J)}] \in \mathbb{R}^{T \times J}$. Hadamardov produkt za dvije matrice jednakih dimenzija je produkt po elementima označen se sa $\odot$, i definiran je kao $(\mathbf{A} \odot \mathbf{B})_{ij} = a_{ij}b_{ij}$. Svojstva navedenih produkata se mogu pronaći u [1].

Produkt tenzora $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ i matrice $\mathbf{A} \in \mathbb{R}^{J_n \times I_n}$ po modu n označava se sa

$$\underline{\mathbf{Y}} = \underline{\mathbf{X}} \times_n \mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_{n-1} \times J_n \times I_{n+1} \times \dots \times I_N} \quad (3)$$

i po komponentama je definiran kao

$$y_{i_1 i_2 \dots i_{n-1} j_n i_{n+1} \dots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \dots i_n} a_{j_n i_n}.$$

Često se produkt tenzora i matrice po modu n prikazuje korištenjem matricizacije i matričnog množenja kao

$$\mathbf{Y}_{(n)} = \mathbf{A} \mathbf{X}_{(n)}.$$

što je samo drugačiji zapis izraza (3). Tenzor možemo množiti matricom po više modova i vrijedi

$$(\underline{\mathbf{X}} \times_n \mathbf{A}) \times_m \mathbf{B} = (\underline{\mathbf{X}} \times_m \mathbf{B}) \times_n \mathbf{A} = \underline{\mathbf{X}} \times_n \mathbf{A} \times_m \mathbf{B},$$

za $m \neq n$. U slučaju da množimo uzastopce po istom modu vrijedi $(\underline{\mathbf{X}} \times_n \mathbf{A}) \times_n \mathbf{B} = \underline{\mathbf{X}} \times_n (\mathbf{B}\mathbf{A})$, gdje se podrazumijeva da matrice $\mathbf{A}$ i $\mathbf{B}$ imaju prikladne dimenzije. Detaljniji opis svojstava množenja može se naći u [1].

U primjenama se često susreću posebne vrste tenzora. Ranije smo definirali tenzor ranga jedan kao tenzor reda N koji se može prikazati u obliku (2). Tenzor nazivamo kubičnim ako ima istu dimenziju u svakom modu, odnosno $\underline{\mathbf{X}} \in \mathbb{R}^{I \times I \times \dots \times I}$. Kubični tenzor nazivamo simetričnim ako su mu elementi jednaki za sve permutacije indeksa, tj. ako je $x_{i_1 i_2 \dots i_N} = x_{\pi(i_1 i_2 \dots i_N)}$ za sve $i_1, \dots, i_N \in \{1, \dots, I\}$, gdje $\pi(i_1 i_2 \dots i_N)$ označava proizvoljnu permutaciju skupa indeksa [30]. Za proizvoljni tenzor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ kažemo da je dijagonalan ako mu za elemente vrijedi $x_{i_1 i_2 \dots i_n} \neq 0$ samo ako je $i_1 = i_2 = \dots = i_N$.

Posebno je važan pojam ranga tenzora. Podsjetimo se prvo definicije ranga matrice. Rang po stupcima matrice $\mathbf{A} \in \mathbb{R}^{I \times J}$ je najveći broj njenih linearno nezavisnih stupaca, odnosno dimenzija slike ako matricu promatramo kao linearni operator između vektorskih prostora. Analogno se definira rang po retcima, kao najveći broj linearno nezavisnih redaka matrice $\mathbf{A}$. Za matrice vrijedi da su rang po stupcima i rang po retcima jednaki pa ih zovemo zajedničkim imenom rang matrice $\mathbf{A}$ i označavamo sa $r(\mathbf{A})$, te vrijedi $r(\mathbf{A}) \leq \min\{I, J\}$. Ekvivalentna definicija ranga matrice dana izrazom

$$r(\mathbf{A}) = \min\{r : \mathbf{A} = \sum_{i=1}^r \mathbf{u}_i \mathbf{v}_i^T, \mathbf{u}_i \in \mathbb{R}^I, \mathbf{v}_i \in \mathbb{R}^J\}, \quad (4)$$

odnosno kao najmanji broj matrica ranga 1 takvih da je njihova suma jednaka matrici $\mathbf{A}$. Općenito, prikaz matrice u obliku (4) nije jedinstven. Jedinstven prikaz matrice u obliku sume matrica ranga 1 se može dobiti samo uz dodatna ograničenja na vektore $\mathbf{u}_i$ i $\mathbf{v}_i$. Definiramo k -rang matrice $\mathbf{A}$ kao maksimalni broj k takav da je bilo kojih k stupaca te matrice linearno nezavisno i označava se sa $k(\mathbf{A})$. Očito vrijedi $k(\mathbf{A}) \leq r(\mathbf{A})$. Matrica $\mathbf{A}$ ima nul-stupac ako i samo ako je $k(\mathbf{A}) = 0$ i $\mathbf{A}$ sadrži proporcionalne (kolinearne) stupce ako je $k(\mathbf{A}) = 1$ [31].

Rang tenzora $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ definiran je kao najmanji broj tenzora ranga jedan čija je suma jednaka početnom tenzoru [2], [32], odnosno izrazom

$$r_*(\underline{\mathbf{X}}) = \min\{r : \underline{\mathbf{X}} = \sum_{i=1}^r \mathbf{a}_i^{(1)} \circ \mathbf{a}_i^{(2)} \circ \dots \circ \mathbf{a}_i^{(N)}\} \quad (5)$$

Ovako definiran rang označavati ćemo sa $r_*(\underline{\mathbf{X}})$. Iako je ova definicija ranga tenzora analogna definiciji ranga matrice (4), svojstva im se bitno razlikuju. Primjerice, za razliku od matričnog ranga ne postoji jednostavan algoritam za određivanje ranga tenzora. Štoviše, taj je problem NP-težak [33], [34]. Za razliku od matričnog slučaja, rang tenzora višeg reda nije odozgo ograničen dimenzijama u pojedinom od modova. Također, rang tenzora čiji su elementi slučajni brojevi ne dostiže nužno teoretski maksimalni rang za tenzor određene dimenzije [12]. Promatranjem tenzora čiji elementi su slučajni brojevi iz neke neprekinute distribucije definira se tipični rang i generički rang. Tipični rang tenzora je svaki onaj rang koji se pojavljuje uz vjerojatnost veću od nule. Ako postoji samo jedan tipični rang koji se pojavljuje uz vjerojatnost 1 tada ga nazivamo generičkim rangom [27], [28].

Postoje i druge definicije ranga za tenzore. U analizi tenzora često se koristi multilinearni rang koji proizlazi iz interpretacije tenzora kao linearnog operatora. Općenito, tenzor $\underline{\mathbf{X}}$ reda N definira $2^N - 2$ netrivialnih linearnih operatora [28]. Posebno, linearni operator sa prostora $\circ_{k=1, i \neq n}^N \mathbb{V}^{(k)}$ u prostor $\mathbb{V}^{(n)}$ definiran je upravo matricom $\mathbf{X}_{(n)}$, odnosno matricizacijom tenzora $\underline{\mathbf{X}}$ po modu n . Tada rang u modu n tenzora $\underline{\mathbf{X}}$ definiramo kao rang matrice $\mathbf{X}_{(n)}$

i označavamo sa $r_n(\mathbf{X})$. Multilinearni rang tenzora reda N definira se kao N -torka koja se sastoji od rangova svakog od modova i označava se sa

$$r_{\boxplus}(\mathbf{X}) = (r_1(\mathbf{X}), r_2(\mathbf{X}), \dots, r_N(\mathbf{X})).$$

Definicija multilinearog ranga analogna je definiciji ranga po stupcima i ranga po retcima za matrice. Kod matrica su rang u modu 1 (rang po retcima) i rang u modu 2 (rang po stupcima) isti i jednaki su rangu definiranom preko (4). Kod tenzora rangovi po modovima u općem slučaju nisu isti. Postoje i druge definicije ranga za tenzore, npr. simetrični rang i nenegativni rang [28], [25], [35].

B. Multilinearni modeli

Pretpostavimo da imamo zadane podatke u formi tenzora $\mathbf{X}$ reda N . Obično u analizi želimo te podatke prikazati korištenjem nekog modela. Korištenjem modela $\hat{\mathbf{X}}$ podatke možemo zapisati kao

$$\mathbf{X} = \hat{\mathbf{X}} + \mathbf{E},$$

odnosno vrijedi

$$\mathbf{X} \approx \hat{\mathbf{X}}.$$

U tom slučaju $\hat{\mathbf{X}}$ predstavlja strukturni dio podataka koji je opisan odabranim modelom, dok je sa $\mathbf{E}$ opisan rezidual odnosno dio varijacije unutar podataka koji odabrani model ne može opisati. Obično se model $\hat{\mathbf{X}}$ sastoji od niza faktora koje koristimo za koncizan opis informacije u izvornim podacima. Kod bilinearnih ili multilinearne modela faktori su linearne kombinacije izvornih podataka. U analizi višedimenzionalnih podataka često su korištene matrične tehnike, primjerice analiza nezavisnih komponenti (eng. *Independent Component Analysis*, ICA), faktorska analiza (eng. *Factor Analysis*, FA) ili faktORIZACIJA nenegativnih matrica (eng. *Nonnegative Matrix Factorization*), međutim to može dovesti do gubitka informacije ili krive interpretacije faktora [1], [24]. Dva najvažnija multilinearne modela koja se koriste u praksi su kanonska poliadska dekompozicija (CPD) i Tuckerov model koji su predstavljeni u nastavku.

Zbog jednostavnosti neki su rezultati predstavljeni za tenzore reda 3, iako vrijede u općem slučaju. U ostatku sekcije $\mathbf{X}$ označava tenzor reda 3, odnosno $\mathbf{X} \in \mathbb{R}^{I \times J \times K}$, ako nije drugačije navedeno.

C. CPD model

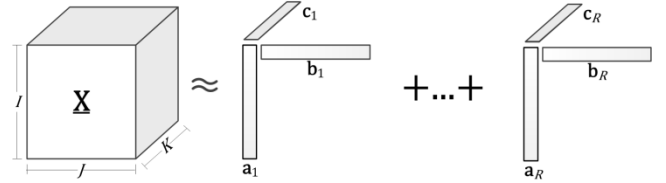
Prvo predstavljanje modela kanonske poliadske dekompozicije (CPD) je u radovima Hitchcocka [2], [3] i Catella [4] ali je postao šire poznat tek nakon radova Harshmana (pod nazivom PARAFAC) [9] i Carolla i Changa (pod nazivom CANDECOMP) [10]. CPD model dekomponira tenzor u sumu tenzora ranga jedan. Za tenzor $\mathbf{X} \in \mathbb{R}^{I \times J \times K}$ reda 3 CPD model zapisujemo kao

$$\mathbf{X} \approx \hat{\mathbf{X}} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \quad (6)$$

gdje je R pozitivan cijeli broj, $\mathbf{a}_r \in \mathbb{R}^I$, $\mathbf{b}_r \in \mathbb{R}^J$ i $\mathbf{c}_r \in \mathbb{R}^K$ za $r \in \{1, \dots, R\}$. Na razini elemenata CPD model može se zapisati kao

$$x_{i,j,k} \approx \sum_{r=1}^R a_{ir} b_{jr} c_{kr},$$

za sve $i \in \{1, \dots, I\}$, $j \in \{1, \dots, J\}$ i $k \in \{1, \dots, K\}$. CPD model je ilustriran na slici Slika 2.



Slika 2: Ilustracija CPD modela za tenzor reda 3.

Vektore iz svakog od modova možemo grupirati u matrice $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_R]$, $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_R]$ i $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_R]$ koje nazivamo faktorskim matricama dok same vektore zovemo faktorima. Korištenjem produkta tenzora i matrice CPD model se može zapisati kao

$$\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{I} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}, \quad (7)$$

gdje je $\mathbf{I} \in \mathbb{R}^{R \times R \times R}$ dijagonalni tenzor reda 3 sa jedinicama na dijagonali. Također je koristan prikaz CPD modela u matriciziranoj formi za svaki od 3 moda

$$\mathbf{X}_{(1)} \approx \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T,$$

$$\mathbf{X}_{(2)} \approx \mathbf{B}(\mathbf{C} \odot \mathbf{A})^T,$$

$$\mathbf{X}_{(3)} \approx \mathbf{C}(\mathbf{B} \odot \mathbf{A})^T.$$

Za tenzor reda 3 ponekad je prikladno zapisati CPD model preko frontalnih odsječaka kao

$$\mathbf{X}_k \approx \hat{\mathbf{X}}_k = \mathbf{A} \mathbf{D}^{(k)} \mathbf{B}^T,$$

gdje su $\mathbf{X}_k$, $\hat{\mathbf{X}}_k$ frontalni odsječci pripadnih tenzora, indeksirani sa k , a $\mathbf{D}^{(k)} = \text{diag}(\mathbf{c}_k)$ dijagonalna matrica koja na dijagonali ima k -ti redak matrice $\mathbf{C}$, za $k \in \{1, \dots, K\}$. Iz navedenih prikaza jasno je da je model ograničen u smislu da i -ti faktor iz jednog moda može međudjelovati samo sa i -tim faktorom iz ostalih modova, te zajedno tvore tenzor ranga jedan.

Ideja CPD modela je aproksimirati tenzor $\mathbf{X}$ modelom, odnosno tenzorom, ranga R . Drugim riječima želimo strukturni dio varijacije unutar podataka prikazati kao sumu R tenzora ranga jedan, kao u (6). U slučaju da je tenzor $\mathbf{X}$ upravo ranga R vrijedi jednakost

$$\mathbf{X} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r. \quad (8)$$

Kažemo da je (8) dekompozicija ranga (eng. *rank revealing decomposition*). Interesantno svojstvo tenzora višeg reda i ranga R je da je njihova dekompozicija u obliku (8) često

jedinstvena. Primijetimo da u matričnom slučaju prikaz proizvoljne matrice kao sume matrica ranga 1 nije jedinstven bez dodatnih ograničenja. Neka je $\mathbf{A} \in \mathbb{R}^{I \times J}$ matrica čija dekompozicija ranga ima oblik

$$\mathbf{A} = \mathbf{U}\mathbf{V}^T = \sum_{r=1}^R \mathbf{u}_r \circ \mathbf{v}_r.$$

Tada možemo uzeti bilo koju ortogonalnu matricu $\mathbf{W} \in \mathbb{R}^{R \times R}$ i konstruirati drugu dekompoziciju kao $\mathbf{A} = (\mathbf{U}\mathbf{W})(\mathbf{V}\mathbf{W})^T$. Dekompozicija ranga za matricu jedinstvena je samo u slučaju dodatnih ograničenja, kao npr. ortogonalnost (SVD dekompozicija).

S druge strane CPD dekompozicija (8) jedinstvena je do na permutaciju i skaliranje stupaca faktorskih matrica uz puno slabije uvjete. Najpoznatiji dovoljan uvjet za jedinstvenost CPD dekompozicije (8) predstavljen u Kruskalovim radovima [32], [36] je

$$k(\mathbf{A}) + k(\mathbf{B}) + k(\mathbf{C}) \geq 2R + 2.$$

Općenito, za tenzor $\underline{\mathbf{X}}$ reda N i ranga R imamo CPD dekompoziciju

$$\underline{\mathbf{X}} = \sum_{r=1}^R \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \dots \circ \mathbf{a}_r^{(N)},$$

a dovoljan uvjet za jedinstvenost [37] dan je sa

$$\sum_{n=1}^N k(\mathbf{A}^{(n)}) \geq 2R + (N - 1). \quad (9)$$

Navedeni uvjet ne može biti zadovoljen za $R = 1$, ali za taj slučaj jedinstvenost je dokazao Harshman [38]. U radu [39] pokazano je da su navedeni uvjeti ujedno i dovoljni za $R = 2$ i $R = 3$. U općem slučaju, nužan uvjet za jedinstvenost je, [40],

$$\min_{n=1, \dots, N} r(\mathbf{A}^{(1)} \odot \dots \odot \mathbf{A}^{(n-1)} \odot \mathbf{A}^{(n+1)} \odot \dots \odot \mathbf{A}^{(N)}) = R.$$

Kako vrijedi $r(\mathbf{A} \odot \mathbf{B}) \leq r(\mathbf{A} \otimes \mathbf{B}) \leq r(\mathbf{A}) \cdot r(\mathbf{B})$, može se dobiti još jednostavniji nužan uvjet za jedinstvenost CPD dekompozicije

$$\min_{n=1, \dots, N} \left(\prod_{m \neq n}^N r(\mathbf{A}^{(m)}) \right) \geq R.$$

U praksi, kada realne podatke opisujemo CPD modelom, pretpostavljamo da je rang strukturnog dijela varijacije (ponekad se naziva pseudo-rank) relativno mali u odnosu na stvarni rang podatkovnog tenzora (tj. njegov tipični rang) [41]. Spomenimo samo da je problem aproksimacije tenzora ranga R drugim tenzorom ranga $k < R$ (eng. *low rank approximation*) bitno složeniji nego ekvivalentni problem u matričnom okruženju. Za detaljniju diskusiju pogledati [25], [28], [27], [35], [42] i reference u navedenim radovima.

Algoritmi za dekompoziciju.

Pretpostavimo da imamo tenzor $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$ reda 3 i nepoznatog ranga. Od ranije je poznato da ne postoji efikasan algoritam za određivanje ranga tenzora [33], [34], pa se postavlja pitanje određivanja broja faktora za CPD model. U praksi se navedeni problem većinom rješava heuristikama. Primjerice, radi se dekompozicija izvornog tenzora u CPD modele sa različitim brojem komponenti i odabere se onaj model koji je najbolji po nekom kriteriju. Najčešće se kao kriterij bira neka funkcija reziduala, npr. norma, ali u slučaju kada je u podacima prisutan veći udio šuma sami reziduali nisu pogodan kriterij. U [43] je predstavljena metoda za određivanje broja komponenti pod nazivom CORCONDIA koja koristi činjenicu da se u (7) pri idealnoj CPD dekompoziciji nalazi dijagonalni tenzor. Također, moguće je koristiti DIFFIT metodu [44], [45] ili heuristiku baziranu na konveksnoj ljusci [46]. U principu ne postoji optimalna metoda za određivanje broja faktora u modelu nego se u kontekstu određenog problema obično koristi konsenzus više metoda.

Neka je broj faktora određen i jednak R . Tada se problem dekompozicije tenzora $\underline{\mathbf{X}}$ u CPD model svodi na traženje tenzora (8) ranga R koji najbolje aproksimira $\underline{\mathbf{X}}$, odnosno na sljedeći problem

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} D(\underline{\mathbf{X}}, \underline{\hat{\mathbf{X}}}), \text{ t.d. } \underline{\hat{\mathbf{X}}} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r.$$

Ciljna funkcija $D(\cdot, \cdot)$ predstavlja mjeru blizine između dva tenzora, čime kvantificiramo kvalitetu aproksimacije izvornog tenzora $\underline{\mathbf{X}}$ pomoću aproksimacije $\underline{\hat{\mathbf{X}}}$. Najčešće se za ciljnu funkciju uzima norma razlike između izvornog tenzora i aproksimacije, te se traži rješenje problema

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \|\underline{\mathbf{X}} - \underline{\hat{\mathbf{X}}}\| \quad (10)$$

što predstavlja nelinearnu verziju problema najmanjih kvadrata [47], [48]. Važno je napomenuti da ciljna funkcija u (10) nije konveksna, što rezultira optimizacijskim problemom sa lokalnim minimumima.

Najčešće korišten pristup za rješavanje problema (10) je metoda alternirajućih najmanjih kvadrata (eng. *Alternating Least Squares*, ALS) koju primjenjujemo tako da pretpostavimo da su u (10) sve osim jedne matrice fiksirane. Time se problem pojednostavljuje na linearni problem najmanjih kvadrata. Ako pretpostavimo da su matrice $\mathbf{B}$ i $\mathbf{C}$ fiksirane (10) se može zapisati kao

$$\min_{\mathbf{A}} \|\mathbf{X}_{(1)} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T\|_F. \quad (11)$$

Kako imamo samo minimizaciju po $\mathbf{A}$ to je problem najmanjih kvadrata sa eksplicitnim rješenjem koje se može zapisati kao

$$\mathbf{A} = \mathbf{X}_{(1)}[(\mathbf{C} \odot \mathbf{B})^T]^\dagger,$$

gdje $\dagger$ označava Moore-Penrose pseudoinverz matrice. Korištenjem svojstava Khatri-Rao produkta rješenje se može jednostavnije zapisati kao

$$\mathbf{A} = \mathbf{X}_{(1)}(\mathbf{C} \odot \mathbf{B})(\mathbf{C}^T \mathbf{C} \odot \mathbf{B}^T \mathbf{B})^\dagger.$$

Za matrice $\mathbf{B}$ i $\mathbf{C}$ dobivamo analognim postupkom izraze

$$\begin{aligned}\mathbf{B} &= \mathbf{X}_{(2)}(\mathbf{C} \odot \mathbf{A})(\mathbf{C}^T \mathbf{C} \odot \mathbf{A}^T \mathbf{A})^\dagger, \\ \mathbf{C} &= \mathbf{X}_{(3)}(\mathbf{B} \odot \mathbf{A})(\mathbf{B}^T \mathbf{B} \odot \mathbf{A}^T \mathbf{A})^\dagger.\end{aligned}$$

Iterativno ponavljanje proceduru estimacije svake od matrice daje algoritam koji se obično naziva CP-ALS ili PARAFAC-ALS. Cijeli algoritam za slučaj tenzora reda N prikazan je u Algoritam 1.

Algoritam 1: CP-ALS dekompozicija

CP-ALS($\underline{\mathbf{X}}, R$)
inicijaliziraj $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$, $n \in \{1, \dots, N\}$
repeat
 for $n=1:N$
 $\mathbf{V} = \bigoplus_{i=1, i \neq n}^N (\mathbf{A}^{(i)T} \mathbf{A}^{(i)})$
 $\mathbf{A}^{(n)} = \mathbf{X}_{(n)}(\bigoplus_{i=1, i \neq n}^N \mathbf{A}^{(n)}) \mathbf{V}^\dagger$
 end
until(konvergencija)
return($\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(n)}$)

Faktorske matrice se mogu inicijalizirati korištenjem lijevih singularnih vektora matricizacije izvornog tenzora po pripadnom modu ili slučajno. Izvođenje iteracija prekidamo kada nema bitnog poboljšanja u ciljnoj funkciji u uzastopnim iteracijama, nema promjene faktorskih matrica, ciljna funkcija je blizu nule ili smo dostigli maksimalni broj iteracija. Algoritam je iznimno jednostavan za implementaciju i nije računski zahtjevan, ali uz te prednosti dolaze i neki nedostaci. Često je potrebno puno iteracija za konvergenciju, a pri tome algoritam ne mora konvergirati globalnom pa čak ni lokalnom optimumu izvornog problema (10). U uzastopnim iteracijama ciljna funkcija sigurno ne raste, ali najbolje što možemo pouzdano reći je da se algoritam zaustavlja kada ciljna funkcija prestaje opadati. Također, konačno rješenje može bitno ovisiti o inicijalizaciji. Razni pristupi poboljšavaju konvergenciju i numerička svojstva CP-ALS algoritma korištenjem naprednijih optimizacijskih tehnika i regularizacije [25]. U praksi se CP-ALS pokazuje kao iznimno koristan i efikasan algoritam, te po kvaliteti rješenja kompetitivan naprednijim metodama za CPD dekompoziciju u usporedbi na simuliranim i stvarnim podacima [49], [41].

Druga klasa algoritama za CPD dekompoziciju rješava problem (10) korištenjem metoda za nelinearne najmanje kvadrate. U tom slučaju se ekvivalentna ciljna funkcija može zapisati u obliku

$$\gamma(\mathbf{p}) = \|\mathbf{X}_{(1)} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T\|_F^2, \quad (12)$$

gdje je $\mathbf{p} = [(\text{vec} \mathbf{A})^T, (\text{vec} \mathbf{B})^T, (\text{vec} \mathbf{C})^T]^T$ [27]. Za razliku od ALS pristupa u ovom slučaju se ciljna funkcija promatra kao funkcija svih parametara CPD modela istovremeno, odnosno gledamo utjecaj elemenata svih faktorskih matrica. Za minimizaciju se može koristiti obična metoda silaska gradijenta, te napredniji pristupi kao što su Gauss-Newton, Levenberg-Marquard (prigušeni (eng. *damped*) Gauss-Newton) ili nelinearni konjugirani gradijent [25], [27], [49], [41], [50].

U praktičnim primjenama CPD modela ipak se ponekad postavljaju određeni zahtjevi na faktorske matrice, primjerice nenegativnost, koji omogućavaju bolju interpretaciju dekompozicije u kontekstu određenog problema [1].

D. Tucker model

Tuckerova dekompozicija je općenitiji model koji uključuje CPD kao poseban slučaj. Izvorno je predstavljena u radovima [6], [7], [8] za slučaj tenzora reda 3 i može se smatrati generalizacijom analize glavnih komponenti i dekompozicije singularnih vrijednosti za tenzore višeg reda. Tenzor $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$ reda 3 dekomponira se prema Tucker modelu kao

$$\underline{\mathbf{X}} \approx \underline{\hat{\mathbf{X}}} = \underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}, \quad (13)$$

gdje je $\underline{\mathbf{G}} \in \mathbb{R}^{P \times Q \times R}$ jezgri tenzor (eng. *core tensor*) ili kraće jezgra, a matrice $\mathbf{A} \in \mathbb{R}^{I \times P}$, $\mathbf{B} \in \mathbb{R}^{J \times Q}$, $\mathbf{C} \in \mathbb{R}^{K \times R}$ su faktorske matrice i njihove stupce nazivamo faktorima. Često se faktorske matrice biraju tako da budu ortogonalne, pa se u tom slučaju ova dekompozicija može gledati kao poopćenje PCA. U Tucker modelu P , Q i R određuju broj faktora u modovima 1, 2 i 3. U slučaju kada je broj faktora manji od dimenzije izvornog tenzora u pripadnom modu jezgri tenzor $\underline{\mathbf{G}}$ možemo smatrati komprimiranom verzijom izvornog tenzora $\underline{\mathbf{X}}$. Kako se Tucker model često koristi za kompresiju, smanjenje dimenzionalnosti te izdvajanje značajki obično je jezgra manje dimenzije od izvornog tenzora, tj. u kontekstu modela (13) vrijedi $P < I$, $Q < J$ i $R < K$ (iako to u općem slučaju nije nužno). Izraz (13) može se prikazati nešto drugačije kao

$$\underline{\mathbf{X}} \approx \underline{\hat{\mathbf{X}}} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} \mathbf{a}_p \mathbf{b}_q \mathbf{c}_r, \quad (14)$$

korištenjem stupaca faktorskih matrica, ili na razini elemenata kao

$$x_{ijk} \approx \hat{x}_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} a_{ip} b_{jq} c_{kr},$$

za sve $i \in \{1, \dots, I\}$, $j \in \{1, \dots, J\}$ i $k \in \{1, \dots, K\}$. Ilustracija Tucker dekompozicije prikazana je na slici Slika 3. Vidi se da elementi jezgre određuju razinu sprege između faktora u pojedinim modovima. Usporedbom izraza (13) i (14) sa (6) i (7) jasno je da se CPD model može promatrati kao restrikcija Tucker modela. U slučaju CPD modela jezgri tenzor je kubičan, i to dijagonalan sa jedinicama na dijagonali. Korisno je model (13) zapisati u matričnom obliku korištenjem matricizacije

$$\begin{aligned}\hat{\mathbf{X}}_{(1)} &= \mathbf{A}\mathbf{G}_{(1)}(\mathbf{C} \otimes \mathbf{B})^T, \\ \hat{\mathbf{X}}_{(2)} &= \mathbf{B}\mathbf{G}_{(2)}(\mathbf{C} \otimes \mathbf{A})^T, \\ \hat{\mathbf{X}}_{(3)} &= \mathbf{C}\mathbf{G}_{(3)}(\mathbf{B} \otimes \mathbf{A})^T.\end{aligned}$$

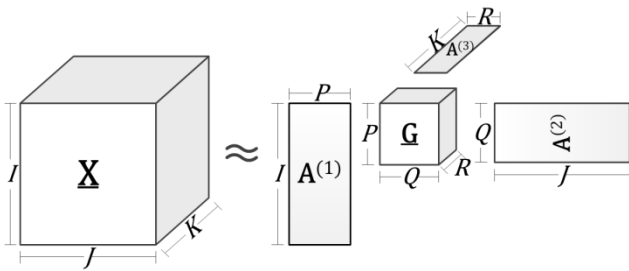
Model (13) se često naziva Tucker3 modelom. Varijanta Tucker modela u kojoj jednu od faktorskih matrica fiksiramo kao jediničnu naziva se Tucker2 model. Primjerice, Tucker2 model sa jediničnom matricom $\mathbf{C} = \mathbf{I} \in \mathbb{R}^{K \times K}$ u modu 3 je

$$\underline{\mathbf{X}} \approx \hat{\underline{\mathbf{X}}} = \underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B}, \quad (15)$$

što može matričnom slučaju zapisati u ekvivalentnom matričnom obliku kao

$$\underline{\mathbf{X}}_k \approx \hat{\underline{\mathbf{X}}}_k = \mathbf{A} \underline{\mathbf{G}}_k \mathbf{B}^T, \text{ za } k \in \{1, \dots, K\}, \quad (16)$$

gdje su $\underline{\mathbf{X}}_k$, $\hat{\underline{\mathbf{X}}}_k$ i $\underline{\mathbf{G}}_k$ frontalni odsječci pripadnih tenzora, indeksirani sa k . Jasno da u ovom slučaju nemamo smanjenje dimenzionalnosti u modu 3, tj. $\underline{\mathbf{G}} \in \mathbb{R}^{P \times Q \times K}$. Slično se dobiva i Tucker1 model, tako da se dvije od 3 faktorske matrice postave u jediničnu.



Slika 3: Ilustracija Tucker3 modela

Analogijom imamo poopćenje za tenzor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ reda N kao

$$\underline{\mathbf{X}} \approx \hat{\underline{\mathbf{X}}} = \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \dots \times_N \mathbf{A}^{(N)}, \quad (17)$$

gdje je $\underline{\mathbf{G}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ i $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$. Izraz (17) možemo prikazati u maticiziranoj formi po modu n kao

$$\mathbf{X}_{(n)} \approx \hat{\mathbf{X}}_{(n)} = \mathbf{A}^{(n)} \underline{\mathbf{G}}_{(n)} (\mathbf{A}^{(N)} \otimes \dots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \dots \otimes \mathbf{A}^{(1)})^T.$$

Model (17) naziva se i Tucker- N model. Postavljanjem jedne od matrica u jediničnu dobije se Tucker- $(N-1)$ model, fiksiranjem dvije Tucker- $(N-2)$, analogno i ostale.

Veća fleksibilnost Tucker modela posljedica je slobode u izboru elemenata jezgre, što donosi više mogućnosti u praktičnim primjenama. Glavni nedostatak fleksibilnosti je što za razliku od CPD modela u ovom slučaju ne možemo jedinstveno odrediti faktorske matrice. Ako uzmemo regularne matrice $\mathbf{U} \in \mathbb{R}^{P \times P}$ tada vrijedi $\hat{\underline{\mathbf{X}}} = \underline{\mathbf{G}} \times_1 \mathbf{A} = \underline{\mathbf{G}}' \times_1 \mathbf{A}'$, gdje je $\underline{\mathbf{G}}' = \underline{\mathbf{G}} \times_1 \mathbf{U}$ nova jezgra, a $\mathbf{A}' = \mathbf{A}\mathbf{U}^{-1}$ nova faktorska matrica. Analognu transformaciju možemo

zbog multilinearnosti primijeniti za faktorske matrice u svim modovima. U primjenama se jedinstvenost pokušava dobiti uvođenjem uvjeta za faktorske matrice i jezgru. Primjerice nenegativnost, rijetkost ili ortogonalnost faktorskih matrica te određena struktura jezgre ograničavaju model [1], [25], [24]. Naravno, tip ograničenja ovisi o primjeni i tipu podataka koji se analizira.

Algoritmi za dekompoziciju.

Prvi problem kod traženja Tucker dekompozicije tenzora je određivanje dimenzija jezgre. U ovom slučaju je taj problem izraženiji nego kod CPD modela jer za tenzor reda N imamo N nepoznatih dimenzija jezgre. Razvijene su razne metode koje koriste maticizacije po pojedinim modovima za određivanje multilinearog ranga [51], [52], [53]. Također su predstavljene i naprednije metode [44], [45], [46], ali ne postoji optimalno rješenje, nego je model potrebno prilagoditi konkretnoj primjeni na bazi više kriterija definiranih problemom koji se rješava. U nastavku ćemo pretpostaviti da su dimenzije jezgre fiksirane i obično manje od pripadne dimenzije izvornog tenzora.

Iako postoji mnogo algoritama za dekompoziciju tenzora uz razne vrste ograničenja u primjenama su posebno česta dva jednostavna algoritma: HOSVD (eng. *Higher-Order Singular Value Decomposition*) i HOOI (eng. *Higher-Order Orthogonal Iteration*). U obje dekompozicije faktorske matrice su ortogonalne.

HOSVD je prvi put predstavljen [8], a detaljno je matematički obrađen u [54] i može se smatrati jednostavnim poopćenjem dekompozicije singularnih vrijednosti na tenzore višeg reda. Ideja je predstaviti tenzor u modu n pomoću komponenti koje optimalno opisuju varijaciju samo u tom modu, neovisno o drugim modovima. Taj pristup nužno vodi na SVD dekompoziciju. U [54] je pokazano da se svaki tenzor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ može prikazati u obliku

$$\underline{\mathbf{X}} = \underline{\mathbf{S}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \dots \times_N \mathbf{U}^{(N)},$$

gdje su matrice $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times I_n}$ ortogonalne, a jezgri tenzor $\underline{\mathbf{S}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ ima svojstvo uređenosti ($\|\underline{\mathbf{S}}_{i_n=1}\| \geq \|\underline{\mathbf{S}}_{i_n=2}\| \geq \dots \geq \|\underline{\mathbf{S}}_{i_n=I_n}\| \geq 0$) i potpune ortogonalnosti ($\langle \underline{\mathbf{S}}_{i_n=\alpha}, \underline{\mathbf{S}}_{i_n=\beta} \rangle = 0$ za $\alpha \neq \beta$). U tom slučaju faktorske matrice $\mathbf{U}^{(n)}$ sadrže lijeve singularne vektore od $\mathbf{X}_{(n)}$, dok su norme podtenzora $\|\underline{\mathbf{S}}_{i_n=i}\|$ jednake i -toj singularnoj vrijednosti od $\mathbf{X}_{(n)}$. HOSVD dekompozicija opisana je u algoritmu Algoritam 2.

Algoritam 2: HOSVD dekompozicija

```

HOSVD( $\underline{\mathbf{X}}$ )
for  $n=1:N$ 
     $\mathbf{U}^{(n)} = \text{lijevi singularni vektori}(\mathbf{X}_{(n)})$ 
end
 $\underline{\mathbf{S}} = \underline{\mathbf{X}} \times_1 \mathbf{U}^{(1)T} \times_2 \mathbf{U}^{(2)T} \dots \times_N \mathbf{U}^{(N)T}$ 
return( $\underline{\mathbf{S}}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(N)}$ )

```

Međutim, u praksi nas često zanima dekompozicija u kojoj jezgra ima manje dimenzije nego izvorni tenzor. U tom slučaju želimo izvorni tenzor $\underline{\mathbf{X}}$ aproksimirati Tucker- N modelom $\hat{\underline{\mathbf{X}}}$ sa multilinearim rangom $r_{\boxplus}(\hat{\underline{\mathbf{X}}}) = (J_1, J_2, \dots, J_N)$ uz uvjet $J_n < r_n(\underline{\mathbf{X}}) \leq I_n$ u (17). Kvaliteta aproksimacije obično se gleda kao norma razlike izvornog tenzora i aproksimacije³, odnosno tražimo model koji je rješenje problema

$$\min_{\underline{\mathbf{G}}, \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}} \|\underline{\mathbf{X}} - \hat{\underline{\mathbf{X}}}\| \quad (18)$$

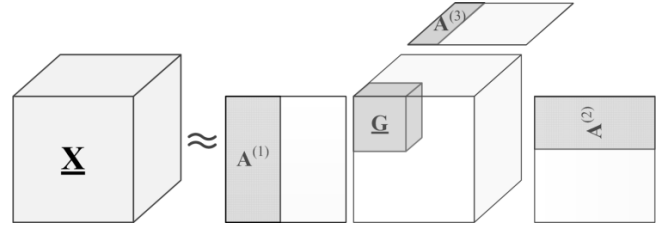
gdje je $r_{\boxplus}(\hat{\underline{\mathbf{X}}}) = (J_1, J_2, \dots, J_N)$. Najjednostavniji način bio bi odbacivanje elemenata u HOSVD dekompoziciji, po uzoru na aproksimaciju nižeg ranga korištenjem SVD dekompozicije u matričnom slučaju. Taj pristup se naziva *truncated HOSVD* (trHOSVD) [54], [25] i ilustriran je na slici Slika 4 za slučaj tenzora reda 3, a u općem slučaju u algoritmu Algoritam 3. Tada vrijedi

$$\underline{\mathbf{X}} \approx \hat{\underline{\mathbf{X}}} = \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \dots \times_N \mathbf{A}^{(N)},$$

gdje $\mathbf{A}^{(n)} = [\mathbf{u}_1^{(n)}, \mathbf{u}_2^{(n)}, \dots, \mathbf{u}_{J_n}^{(n)}] \in \mathbb{R}^{I_n \times J_n}$ sadrže J_n vodećih lijevih singularnih vektora iz moda n . Za razliku između izvornog tenzora i modela vrijedi ocjena

$$\|\underline{\mathbf{X}} - \hat{\underline{\mathbf{X}}}\|^2 \leq \sum_{i_1=J_1+1}^{r_1(\underline{\mathbf{X}})} \sigma_{i_1}^{(1)2} + \sum_{i_2=J_2+1}^{r_2(\underline{\mathbf{X}})} \sigma_{i_2}^{(2)2} + \dots + \sum_{i_N=J_N+1}^{r_N(\underline{\mathbf{X}})} \sigma_{i_N}^{(N)2}$$

gdje je sa $\sigma_{i_k}^{(n)2}$ označena i_k -ta singularna vrijednost u modu n . Iako se na ovaj način postiže dobra aproksimacija, pokazuje se da to nije najbolja aproksimacija u smislu (18) [55].



Slika 4: Tucker3 model uz trHOSVD dekompoziciju

Algoritam 3: HOSVD dekompozicija sa odbacivanjem

```

trHOSVD( $\underline{\mathbf{X}}, J_1, J_2, \dots, J_N$ )
for  $n=1:N$ 
     $\mathbf{A}^{(n)} = J_n \text{ lijevih singularnih vektora}(\mathbf{X}_{(n)})$ 
end
 $\underline{\mathbf{G}} = \underline{\mathbf{X}} \times_1 \mathbf{A}^{(1)T} \times_2 \mathbf{A}^{(2)T} \dots \times_N \mathbf{A}^{(N)T}$ 
return( $\underline{\mathbf{G}}, \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)}$ )

```

Problem najbolje aproksimacije (18) svodi se na traženje ortonormalnih faktorskih matrica $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$, a optimalna jezgra dana je sa $\underline{\mathbf{G}} = \underline{\mathbf{X}} \times_1 \mathbf{A}^{(1)T} \dots \times_N \mathbf{A}^{(N)T}$ [55]. U tom slučaju se problem (18) može formulirati kao

$$\max_{\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}} \|\underline{\mathbf{X}} \times_1 \mathbf{A}^{(1)T} \times_2 \mathbf{A}^{(2)T} \dots \times_N \mathbf{A}^{(N)T}\| \quad (19)$$

gdje su $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ takve da vrijedi $\mathbf{A}^{(n)T} \mathbf{A}^{(n)} = \mathbf{I}$.

Postoje mnogi pristupi za rješavanje problema (19). Najjednostavniji je ALS pristup kojim dobivamo HOOI algoritam [55]. Pretpostavimo da su sve faktorske matrice fiksirane osim faktorske matrice u modu n . U tom slučaju (19) se reducira u jednostavniji optimizacijski problem

$$\max_{\mathbf{A}^{(n)}} \|\mathbf{A}^{(n)T} \mathbf{W}^{(n)}\|_F$$

gdje je

$$\mathbf{W}^{(n)} = \mathbf{X}_{(n)} (\mathbf{A}^{(N)} \otimes \dots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \dots \otimes \mathbf{A}^{(1)}).$$

Rješenje se dobiva SVD dekompozicijom uzimanjem J_n vodećih lijevih singularnih vektora od $\mathbf{W}^{(n)}$. Kao i kod CP-ALS algoritma, iteracije nužno ne moraju konvergirati globalnom ni lokalnom minimumu, nego samo točki infleksije. Cijeli HOOI algoritam je prikazan u algoritmu 4.

³ Općenito se mogu koristiti i druge mjere sličnosti između tenzora, primjerice α ili β divergencije [1], koje vode i na druge algoritme za dekompoziciju.

Algoritam 4: HOOI algoritam

```

HOOI( $\mathbf{X}, J_1, J_2, \dots, J_N$ )
inicijaliziraj  $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}, n \in \{1, \dots, N\}$ 
repeat
  for  $n=1:N$ 
     $\mathbf{W}^{(n)} = \mathbf{X}_{(n)} (\otimes_{i=N, i \neq n}^1 \mathbf{A}^{(i)})$ 
     $\mathbf{A}^{(n)} = J_n$  lijevih sing. vekt. od  $\mathbf{W}^{(n)}$ 
  end
until(konvergencija)
 $\mathbf{G} = \mathbf{X} \times_1 \mathbf{A}^{(1)T} \times_2 \mathbf{A}^{(2)T} \dots \times_N \mathbf{A}^{(N)T}$ 
return( $\mathbf{G}, \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)}$ )

```

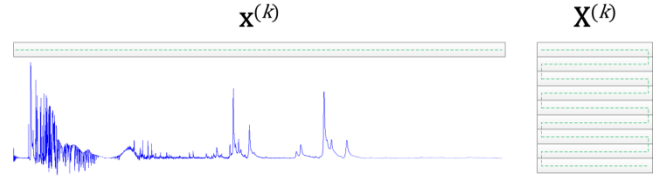
Napredniji algoritam koji osigurava konvergenciju u stacionarnu točku ciljne funkcije predstavljen je u [56]. Za optimizacijski problem (19) koristi se Newtonova metoda na produktu Grassmannovih mnogostrukosti. Algoritam pokazuje kvadratičnu brzinu konvergencije, uz nešto sporije iteracije zbog računanja Hessiana. Metoda je ubrzana korištenjem kvazi-Newton metode i BFGS pristupa u [57]. S druge strane, opširan opis algoritama koji koriste nenegativnost i rijetkost može se naći u [1], [58].

III. PRIMJENE U IZDVAJANJU ZNAČAJKI

Dekompozicije tenzora primjenjivane su u raznim scenarijima za izdvajanje značajki, primjerice u prepoznavanju lica i tekstura [14], [58], pokreta iz slike [59], [60], slika objekata [15], [61], pokreta iz EEG signala [15], [19]. Ovdje ćemo predstaviti primjenu u na problemima klasifikacije bolesti iz proteinskih i genskih profila i klasifikaciji melanoma.

A. Izdvajanje značajki iz proteinskih profila za klasifikaciju bolesti

Proteinski profili dobiveni spektrometrijom mase korišteni su u dijagnostici raznih bolesti, primjerice raka prostate i jajnika [62]. Jedan uzorak (proteinski profil) obično sadrži veliki broj intenziteta za različite omjere mase i naboja m/z koji predstavljaju varijable, tipično 5000 do 20000. Svaki uzorak odgovara jednom pacijentu pa je broj uzoraka relativno mali, čak za nekoliko redova veličine manji od broja varijabli. Zbog toga je potrebno napraviti smanjenje dimenzionalnosti prije klasifikacije, da bi dobili dobru klasifikaciju korištenjem manjeg broja diskriminativnih značajki. Iako su dekompozicije tenzora prikladne za analizu višedimenzionalnih podataka u [63] je predstavljena metoda za izdvajanje značajki iz 1D proteinskih profila. U tu svrhu svaki proteinski profil transformiran je u matricu, na način prikazan na slici Slika 5. Često se u analizi slike korištenjem bilinearnih metoda slike vektoriziraju, dok je ovdje korišten obrnut postupak: svaki proteinski profil (vektor) presložen je u dvodimenzionalni uzorak (matricu).



Slika 5: Prebacivanje 1D uzorka $\mathbf{x}^{(k)}$ u 2D uzorak $\mathbf{X}^{(k)}$

Promatramo problem binarne klasifikacije gdje jedna klasa odgovara zdravim pacijentima (klasa -1), a druga klasa oboljelim pacijentima (klasa 1). Pretpostavimo da se trening skup sastoji od proteinskih profila $\mathbf{x}^{(k)} \in \mathbb{R}^{I_1 \cdot I_2}$ sa pripadnim labelama $c^{(k)} \in \{-1, 1\}$, $k \in \{1, \dots, K\}$, za K pacijenata. Za svaki uzorak radimo matricizaciju čime dobivamo 2D uzorke $\mathbf{X}^{(k)} \in \mathbb{R}^{I_1 \times I_2}$ sa pripadnim labelama $c^{(k)}$. Da bi za svaki uzorak izdvojili značajke promatramo problem zajedničke faktorizacije [15]

$$\mathbf{X}^{(k)} \approx \mathbf{A} \mathbf{F}^{(k)} \mathbf{B}^T, \text{ za } k \in \{1, \dots, K\},$$

gdje su matrice $\mathbf{A} \in \mathbb{R}^{I_1 \times J}$ i $\mathbf{B} \in \mathbb{R}^{I_2 \times J}$ zajednički faktori, a matrica $\mathbf{F}^{(k)} \in \mathbb{R}^{J \times J}$ predstavlja dobivene značajke za k -ti uzorak. Na ovaj način uzorke sa $I_1 \cdot I_2$ varijabli prikazujemo sa J^2 značajki, i obično vrijedi $J^2 \ll I_1 \cdot I_2$. Zbog jednostavnosti smo pretpostavili da je matrica $\mathbf{F}^{(k)}$ kvadratna. Iz (15) i (16) slijedi da zajedničku faktorizaciju možemo zapisati kao Tucker-2 model

$$\mathbf{X} \approx \mathbf{F} \times_1 \mathbf{A} \times_2 \mathbf{B},$$

gdje je $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2 \times K}$ tenzor reda 3 čiji frontalni odsječci odgovaraju uzorcima, odnosno $\mathbf{X}_k = \mathbf{X}^{(k)}$. Analogno, značajke su frontalni odsječci tenzora $\mathbf{G} \in \mathbb{R}^{J \times J \times K}$, odnosno $\mathbf{F}^{(k)} = \mathbf{F}_k$. Za dekompoziciju koristimo HOOI algoritam, iako je moguće koristiti i neke druge. Vektorizirane značajke zajedno sa njihovim labelama $(\mathbf{f}^{(k)}, c^{(k)})$, gdje je $\mathbf{f}^{(k)} = \text{vec}(\mathbf{F}^{(k)})$, koristimo za treniranje odabranog klasifikatora. Da bi mogli procijeniti točnost klasifikacije potrebne su nam značajke za test uzorke, koje dobivamo korištenjem matrica $\mathbf{A}$ i $\mathbf{B}$ naučenih na trening skupu

$$\mathbf{F}_{\text{test}} \approx \mathbf{X}_{\text{test}} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T,$$

gdje je tenzor $\mathbf{X}_{\text{test}} \in \mathbb{R}^{I_1 \times I_2 \times M}$ sadrži M test uzoraka kao frontalne odsječke. Navedeni pristup primijenjen je na problemima klasifikacije pacijenata oboljelih od raka prostate i raka jajnika iz proteinskih profila dobivenih spektrometrijom mase uzoraka proteina u krvnom serumu [64]. Oba skupa podataka sastoji se od dvije klase: (i) zdravi pacijenti (kontrolni uzorci); (ii) pacijenti kod kojih je rak potvrđen biopsijom. Za rak prostate dostupno je 69 uzoraka bolesnih pacijenata i 63 kontrolna uzorka, dok je za rak jajnika dostupno po 100 uzoraka u svakoj klasi. Svaki proteinski profil je duljine 15154, pa je odbačeno zadnjih 25 elemenata prije preslagivanja u matricu dimenzija $I_1 = I_2 = 123$. Za klasifikaciju su korišteni linearni i nelinearni SVM i kNN klasifikator, a točnost klasifikacije je procijenjena dvostrukom krosvalidacijom

uz 200 ponavljanja. Rezultati su prikazani u tablici 1 i usporedivi su ili bolji nego rezultati prijavljeni u literaturi [63].

B. Izdvajanje značajki za klasifikaciju melanoma

Trenutno korištene metode za klasifikaciju kožnih lezija koriste barem jedan od tri postupka: (i) ispitivanje kliničke slike kožne lezije kod iskusnog dermatologa; (ii) ispitivanje morfoloških struktura kožne lezije mikroskopijom kože; (iii) ispitivanje spektralnih svojstava kožne lezije korištenjem hiperspektralne kamere. U navedenim slučajevima potreban je ili visokoobrazovan i iskusen stručnjak ili skup optički sistem da bi karakterizirali kožnu leziju. U svrhu smanjenja broja nepotrebnih biopsija radi se na metodama za automatsku klasifikaciju kožnih lezija koje bi pružile drugo mišljenje dermatologu. U tu svrhu se koriste razne metode za analizu slike dobivene dermoskopijom ili hiperspektralnim sustavima.

Tablica 1: Rezultati klasifikacije za proteinske profile

	J^2	kNN	linSVM	polySVM	rbfSVM
Rak prostate	36	85.4±7.1 / 85.1±5.6 $k=4$	91.9±5.6 / 91.4±4.8	94.4±4.7 / 89.0±5.9 $d=3$	93.2±5.1 / 89.7±5.1 $\sigma=3.2$
	100	89.3±5.4 / 86.7±5.3 $k=2$	98.2±2.8 / 95.6±3.6	98.3±2.7 / 89.2±5.5 $d=3$	94.6±5.2 / 90.3±4.7 $\sigma=6.6$
	256	90.2±5.4 / 89.7±5.1 $k=4$	98.8±2.2 / 96.5±3.4	97.0±3.3 / 90.0±5.4 $d=2$	96.2±3.9 / 91.0±4.9 $\sigma=12.6$
	400	92.6±4.6 / 90.1±4.2 $k=2$	99.3±1.6 / 97.8±3.3	98.2±2.7 / 90.7±5.9 $d=2$	96.4±4.3 / 91.5±4.8 $\sigma=13.2$
	625	93.0±4.9 / 91.8±4.6 $k=2$	99.6±1.2 / 98.7±2.9	98.9±2.1 / 91.4±5.1 $d=2$	98.0±3.2 / 93.2±4.5 $\sigma=19$
Rak jajnika	36	66.7±8.4 / 64.7±7.2 $k=12$	84.3±6.1 / 81.4±5.9	83.4±6.5 / 79.8±7.2 $d=3$	84.6±5.6 / 77.3±6.7 $\sigma=4.4$
	100	77.0±7.0 / 64.6±6.7 $k=4$	91.1±4.6 / 87.7±5.0	90.7±4.5 / 82.7±6.4 $d=3$	92.1±4.2 / 84.5±5.4 $\sigma=8.2$
	256	82.7±6.2 / 67.7±6.5 $k=4$	93.6±3.7 / 91.9±4.1	93.6±3.5 / 85.1±5.8 $d=3$	93.9±3.0 / 87.7±4.8 $\sigma=14.2$
	400	85.4±5.7 / 67.2±7.3 $k=4$	95.5±3.2 / 93.4±3.6	93.4±3.6 / 85.4±5.5 $d=2$	95.1±2.7 / 90.1±4.4 $\sigma=17.4$
	625	87.8±5.3 / 68.7±6.2 $k=4$	96.8±2.9 / 95.4±3.5	94.2±3.4 / 87.1±5.7 $d=2$	95.0±2.7 / 91.9±4.1 $\sigma=18.6$

Osjetljivost i specifičnost u % (srednja vrijednost ± standardno odstupanje) procijenjene dvostrukom krosvalidacijom. Optimalne vrijednosti parametara klasifikatora su dane u tablici (k -broj susjeda za kNN, d -stupanj polinoma za polinomski SVM, σ -širina Gaussove jezgre za rbfSVM za Gaussovom jezgrom).

U idealnom slučaju automatska dijagnostika bi se trebala raditi korištenjem što jednostavnijeg sustava za akviziciju slike. Ideja je iskoristiti obične slike u boji (RGB slike) kao preliminarni korak u klasifikaciji kožne lezije. Kako se RGB slika može predstaviti tenzorom reda 3 mogu se isprobati dekompozicije tenzora kao metoda za izdvajanje značajki. Prvi korak za automatske metode je problem

određivanja je li lezija benigna ili maligna, odnosno binarna klasifikacija malignih melanoma i benignih nevusa.

Neka imamo skup podataka koji se sastoji od K uzoraka (RGB slika melanoma ili benignih nevusa). Svaki uzorak možemo predstaviti tenzorom reda 3 kao $\underline{\mathbf{X}}^{(k)} \in \mathbb{R}_{0+}^{I_1(k) \times I_2(k) \times 3}$, za $k \in \{1, \dots, K\}$. Kako je niska spektralna rezolucija ograničavajući faktor kod RGB slike, nelinearnim preslikavanjem povećavamo dimenziju tenzora po modu 3. Preciznije, svaki slikovni element (piksel) tenzora $\underline{\mathbf{X}}^{(k)}$ preslikavamo nekom nelinearnom funkcijom $\phi_D: \mathbb{R}^3 \mapsto \mathbb{R}^D$ u D dimenzionalni vektor, odnosno $\mathbf{x}_{i_1 i_2}^{(k)} \mapsto \phi_D(\mathbf{x}_{i_1 i_2}^{(k)})$. Preslikavanjem svakog piksela dobivamo ponovno tenzor reda 3 koji označavamo sa $\phi_D(\underline{\mathbf{X}}^{(k)}) \in \mathbb{R}_{0+}^{I_1(k) \times I_2(k) \times D}$. Ideja sa nelinearnom preslikavanju temelji se na Coverovu teoremu [65]. Od svih mogućih preslikavanja odabrali smo empirijsko preslikavanje korištenjem jezgre (eng. *empirical kernel map*, EKM) [66]. Preslikavanje definiramo kao

$$\mathbb{R}^3 \ni \mathbf{x} \mapsto \phi_D = \begin{bmatrix} k(\mathbf{v}_1, \mathbf{x}) \\ \vdots \\ k(\mathbf{v}_D, \mathbf{x}) \end{bmatrix} \in \mathbb{R}^D,$$

gdje je $k: \mathbb{R}^3 \times \mathbb{R}^3 \mapsto \mathbb{R}$ realna pozitivno definitna funkcija koju nazivamo jezgra [66], a $\{\mathbf{v}_1, \dots, \mathbf{v}_D\} \subset \mathbb{R}^3$ skup prikladno odabranih vektora. Cijeli postupak izdvajanja značajki prikazan je u algoritmu 5. Nakon nelinearnog preslikavanja tenzor $\phi_D(\underline{\mathbf{X}}^{(k)})$ dekomponiramo u Tucker3 model. Značajke za klasifikaciju dobivamo projekcijom na baze dobivene Tucker3 modelom. Slike mogu općenito biti različitih dimenzija pa se odabirom istih dimenzija jezgre za sve modele dobiva jednak broj značajki za svaku sliku, te ih možemo koristiti za treniranje i testiranje klasifikatora. Za novi uzorak (koji nije u trening skupu) na isti način obavimo nelinearno preslikavanje i dekompoziciju te elemente jezgre interpretiramo kao značajke. Značajke dobivene na ovaj način su elementi jezgrenog tenzora koji povezuje faktore iz svih modova. Stoga ih je moguće interpretirati kao prostorno-spektralne značajke, jer objedinjuju informacije sadržane u prostornim i spektralnim modovima slike.

Algoritam 5: Izdvajanje značajki za klasifikaciju melanoma

Ulaz: uzorak (RGB slika) $\underline{\mathbf{X}}^{(k)} \in \mathbb{R}^{I_1 \times I_2 \times 3}$

Parametri: EKM (σ, D), dimenzije jezgre (J_1, J_2, J_3)

Nelinearno preslikavanje

$$\underline{\mathbf{X}}^{(k)} \mapsto \phi_D(\underline{\mathbf{X}}^{(k)})$$

Tucker3 dekompozicija (trHOSVD/HOOI)

$$\phi_D(\underline{\mathbf{X}}^{(k)}) \approx \underline{\mathbf{G}}^{(k)} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)}$$

Izdvajanje značajki

$$\underline{\mathbf{F}}^{(k)} = \phi_D(\underline{\mathbf{X}}^{(k)}) \times_1 \mathbf{A}^{(1)T} \times_2 \mathbf{A}^{(2)T} \times_3 \mathbf{A}^{(3)T}$$

Vektorizacija

$$\mathbf{f}^{(k)} = \text{vec}(\underline{\mathbf{F}}^{(k)})$$

Izlaz: značajke $\mathbf{f}^{(k)} \in \mathbb{R}^{J_1 \cdot J_2 \cdot J_3}$

Predstavljena metoda testirana je na binarnom klasifikacijskom problemu skupa kliničkih slika melanoma i benignih nevusa dostupnih na arhivama [67], [68], [69], [70]. Ukupno je sakupljeno 90 slika melanoma i 90 slika benignih nevusa. Sa svake slike je izrezan dio oko same lezije, tako da slika većim dijelom sadržava samu leziju. Za nelinearno preslikavanje korištena je Gaussova jezgra, a skup $\{\mathbf{v}_1, \dots, \mathbf{v}_D\}$ dobiven grupiranjem piksela svake slike k -means algoritmom uz $D=40$. Za klasifikaciju su korišteni linearni i nelinearni SVM klasifikatori, dok je točnost klasifikacije procijenjena dvostrukom krosvalidacijom sa 100 ponavljanja. Parametri klasifikatora, jezgre za EKM i dimenzije jezgrenog tenzora odabrane su krosvalidacijom. Dobiveni rezultati su usporedivi sa rezultatima do sada prijavljenima u literaturi i prikazani su u tablici 2 [71].

Tablica 2: Rezultati klasifikacije melanoma

(J_1, J_2, J_3)	linSVM	polySVM	rbfSVM
	20.0±6.2	8.3±4.1	81.8±5.2
(10,12,12)	96.2±3.1	99.4±1.1	87.2±4.1
		$d=2$	$\sigma_C=18$
	18.4±6.7	8.2±4.1	82.1±5.2
(11,12,10)	95.0±3.2	99.8±0.6	86.9±4.2
		$d=2$	$\sigma_C=19$

Osjetljivost i specifičnost u % (srednja vrijednost ± standardno odstupanje) procijenjene dvostrukom krosvalidacijom. Optimalne vrijednosti parametara klasifikatora su dane u tablici (d -stupanj polinoma za polinomski SVM, σ_C -širina Gaussove jezgre za rbfSVM za Gaussovom jezgrom).

LITERATURA

- [1] A. Cichocki, R. Zdunek, A. H. Phan, and S. I. Amari, *Nonnegative Matrix and Tensor Factorizations*. Chichester: John Wiley & Sons, 2009.
- [2] F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," *Journal of Mathematics and Physics*, no. 7, pp. 164-189, 1927.
- [3] F. L. Hitchcock, "Multiple invariants and generalized rank of a p-way matrix or tensor," *Journal of Mathematics and Physics*, vol. 7, pp. 39-79, 1927.
- [4] R. B. Cattell, "Parallel proportional profiles and other principles for determining the choice of factors by rotation," *Psychometrika*, vol. 9, pp. 267-283, 1944.
- [5] R. B. Cattell, "The three basic factor-analytic research designs - their interrelations and derivatives," *Psychological Bulletin*, vol. 49, pp. 499-452, 1952.
- [6] L. R. Tucker, "Implications of factor analysis of three-way matrices for measurement of change," in *Problems in Measuring Change*, C. W. Hariss, Ed.: University of Winsconsin Press, 1963, pp. 122-137.
- [7] L. R. Tucker, "The extension of factor analysis to three-dimensional matrices," in *Contributions to Mathematical Psychology*, H. Gulliksen and N. Frederiksen, Eds. New York: Holt, Rinehardt & Winston, 1964, pp. 110-127.
- [8] L. R. Tucker, "Some mathematical notes on tree-mode factor analysis," *Psychometrika*, vol. 31, pp. 279-311, 1966.
- [9] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis," *UCLA Working Papers in Phonetics*, Vol. 16, No. 1. (1970) Key: citeulike:3806251, vol. 16, pp. 1-84, 1970.
- [10] J. D. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, pp. 283-319, 1970.
- [11] C. J. Appellof and E. R. Davidson, "Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents," *Analytical Chemistry*, vol. 53, pp. 2053-2056, 1981.
- [12] A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis: Applications in the Chemical Sciences*. West Sussex, England: Wiley, 2004.
- [13] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General Tensor Discriminant Analysis and Gabor Features for Gait Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 1700-1715, 2007.
- [14] W. Zhang, Z. Lin, and T. Xiaou, "Tensor linear Laplacian discrimination (TLLD) for feature extraction," *Pattern Recognition*, vol. 42, pp. 1941-1948, 2009.
- [15] A. H. Phan and A. Cichocki, "Tensor decompositions for feature extraction and classification of high dimensional datasets," *IEICE Transaction on Fundamentals*, vol. 1, pp. 37-68, 2010.
- [16] F. Cong et al., "Benefits of Multi-domain Feature of Mismatch Negativity Extracted by Nonnegative Tensor Factorization from Low-density Array EEG," *International Journal of Neural Systems*, 2012.
- [17] F. Cong et al., "Feature Extraction by Nonnegative Tucker Decomposition from EEG Data Including Testing and Training Observations," in *Proceedings of The International Conference on Neural Information Processing (ICONIP) 2012*, Doha, Qatar, 2012.
- [18] D. Wang and S. Kong, "Feature selection from high-order tensorial data via sparse decomposition," *Pattern Recognition Letters*, vol. 33, pp. 1695-1702, 2012.
- [19] A. H. Phan and A. Cichocki, "Extended HALS algorithm for nonnegative Tucker decomposition and its applications for multiway analysis and classification," *Neurocomputing*, vol. 74, pp. 1956-1969, 2011.
- [20] C. A. Andersson and R. Bro, "The N-way toolbox for MATLAB," *Chemometrics and Intelligent Laboratory Systems*, vol. 52, pp. 1-4, 2000, URL:

<http://www.models.life.ku.dk/nwaytoolbox>.

- [21] S. Gourvenec et al., "CuBatch, a MATLAB interface for n-mode data analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 77, pp. 122-180, 2005, URL: <http://www.models.life.ku.dk/CuBatch>.
- [22] B. W. Bader and T. G. Kolda, (2007) MATLAB Tensor Toolbox. [Online]. <http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox/>
- [23] A. H. Phan, (2011) NFEA: Tensor Toolbox for Feature Extraction and Applications. [Online]. <http://www.bsp.brain.riken.jp/~phan/nfea/nfea.html>
- [24] E. Acar and B. Yener, "Unsupervised Multiway Data Analysis: A Literature Survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 1, pp. 6-20, 2009.
- [25] T.W. Kolda and B. W. Bader, "Tensor Decompositions and Applications," *SIAM Review*, vol. 51, no. 3, pp. 455-500, 2009.
- [26] L.-H. Lim and P. Comon, "Multiarray signal processing: Tensor decomposition meets compressed sensing," *Comptes Rendus Mecanique*, no. 338, pp. 311-320, 2010.
- [27] P. Comon, X. Luciani, and A.L.F de Almeida, "Tensor Decompositions, Alternating Least Squares and other Tales," *Journal of Chemometrics*, vol. 23, pp. 393-405, 2009.
- [28] P. Comon, "Tensors versus matrices: usefulness and unexpected problems," in *IEEE Workshop on Statistical Signal Processing*, Cardiff, 2009.
- [29] H. A. L. Kiers, "Towards a standardized notation and terminology in multiway analysis," *Journal of Chemometrics*, vol. 14, no. 3, pp. 105-122, 2000.
- [30] P. Comon, G. Golub, L.-H. Lim, and B. Mourrain, "Symmetric tensors and symmetric tensor rank," SCCM Technical Report 06-02, Stanford University, 2006.
- [31] A. Stegeman and N. D. Sidiropoulos, "On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition," *Linear Algebra and its Applications*, vol. 420, pp. 540-552, 2007.
- [32] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics," *Linear Algebra and its Applications*, no. 18, pp. 95-138, 1977.
- [33] J. Håstad, "Tensor rank is NP-complete," *Journal of Algorithms*, vol. 11, pp. 644-654, 1990.
- [34] C. J. Hillar and L.-H. Lim, "Most Tensor Problems are NP-Hard," *preprint*, 2012.
- [35] P. Comon and L.-H. Lim, "Multilinear algebra in machine learning and signal processing," in *ICIAM Minisymposium on Numerical Linear Algebra*, 2007.
- [36] J. B. Kruskal, "rank, decomposition, and uniqueness for 3-way and N-way arrays," in *Multiway Data Analysis*, R. Coppi and S. Bolasco, Eds. Amsterdam, North-Holland, 1989, pp. 7-18.
- [37] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *Journal of Chemometrics*, vol. 14, pp. 229-239, 2000.
- [38] R. A. Harshman, "Determination and proof of minimum uniqueness conditions for Parafac-1," *UCLA Working Papers in Phonetics*, vol. 22, pp. 111-117, 1972.
- [39] J. M. F. Ten Berge and N. D. Sidiropoulos, "On uniqueness in CANDECOMP/PARAFAC," *Psychometrika*, vol. 67, pp. 399-409, 2002.
- [40] X. Liu and N. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Transactions on Signal Processing*, vol. 49, pp. 2074-2086, 2001.
- [41] G. Tomasi, *Practical and computational aspects in chemometrics data analysis*, Ph.D. Dissertation. Frederiksberg, Denmark: Department of Food Science, The Royal Veterinary and Agricultural University, 2006.
- [42] V. De Silva and L.-H. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, pp. 1084-1127, 2008.
- [43] R. Bro and H. A. L. Kiers, "A new efficient method for determining the number of components in PARAFAC models," *Journal of Chemometrics*, vol. 17, pp. 274-286, 2002.
- [44] M. E. Timmerman and H. A. L. Kiers, "Three-mode principal component analysis: Choosing the numbers of components and sensitivity to local optima," *British Journal of Mathematical and Statistical Psychology*, vol. 53, pp. 1-16, 2000.
- [45] H. A. L. Kiers and A. der Kinderen, "A fast method for choosing the numbers of components in Tucker3 analysis," *British Journal of Mathematical and Statistical Psychology*, vol. 56, pp. 119-125, 2003.
- [46] E. Ceulemans and H. A. L. Kiers, "Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method," *British Journal of Mathematical and Statistical Psychology*, vol. 59, pp. 133-150, 2006.
- [47] K. Madsen, H.B. Nielsen, and O. Tingleff, *Methods for non-linear least squares problems*.: Informatics and Mathematical Modelling, Technical University of Denmark, 2004.
- [48] A. Björck, *Numerical methods for Least Squares Problems*.:

SIAM, 1996.

- [49] G. Tomasi and R. Bro, "A comparison of algorithms for fitting the PARAFAC model," *Computational Statistics & Data Analysis*, vol. 50, pp. 1700-1734, 2006.
- [50] A. H. Phan, P. Tichavsky, and A. Cichocki, "Low complexity damped Gauss-Newton algorithms for CANDECOMP/PARAFAC," *unpublished manuscript*, 2012.
- [51] Z. He, A. Cichocki, and S. Xie, "Efficient method for Tucker3 model selection," *Electronics Letters*, vol. 45, no. 15, pp. 805-806, 2009.
- [52] Ze He, A. Cichocki, S. Xie, and K. Choi, "Detecting the Number of Clusters in n-Way Probabilistic Clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 3006-2021, 2010.
- [53] E. R. Malinowski, "Determination of the number of factors and experimental error," *Analytical Chemistry*, vol. 49, pp. 612-617, 1977.
- [54] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A Multilinear Singular Value Decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253-1278, 2000.
- [55] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-(R1,R2,RN) approximation of higher-order tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324-1342, 2000.
- [56] L. Elden and B. Savas, "A Newton-Grassmann method for computing the best multilinear rank-(r1,r2,r3) approximation of a tensor," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, pp. 248-271, 2009.
- [57] B. Savas and L. H. Lim, "Quasi-Newton methods on Grassmannians and multilinear approximation of tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 32, no. 6, pp. 3352-3393, 2010.
- [58] S. Yan, D. Xu, Q. Yang, L. Zhang, and X. Zhang, H.-J. Tang, "Discriminant Analysis with Tensor Representation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, USA, 2005, pp. 526-532.
- [59] Z.-Z. Yu, C.-C. Jia, W. Pang, C.-Y. Zhang, and L.-H. Zhong, "Tensor Discriminant Analysis With Multiscale Features for Action Modeling and Categorization," *IEEE Signal Processing Letters*, vol. 19, 2012.
- [60] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1700-1715, 2007.
- [61] F. Nie, S. Xiang, Y. Song, and C. Zhang, "Extracting the optimal dimensionality for local tensor discriminant analysis," *Pattern Recognition*, vol. 42, pp. 105-114, 2009.
- [62] I. Kopriva and M. Filipović, "A mixture model with a reference-based automatic selection of components for disease classification from protein and/or gene expression levels," *BMC Bioinformatics*, vol. 12, 2011.
- [63] I. Kopriva and A. Jukić, "Feature extraction for cancer prediction by tensor decomposition of 1D protein expression levels," in *IASTED Conference on Computational Bioscience CompBio 2011*, Cambridge, UK, 2011.
- [64] Center for Cancer Research, National Cancer Institute. [Online]. <http://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.asp>
- [65] T. M. Cover, "Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition," *IEEE Transactions on Electronic Computers*, vol. EC-14, no. 3, pp. 326-334, 1965.
- [66] B. Schölkopf and A. J. Smola, *Learning with kernels*. Cambridge, MA: MIT Press, 2002.
- [67] T. L. Diepgen, G. Yihune, et al. Dermatology Online Atlas. [Online]. <http://www.dermis.net>
- [68] DermAtlas. [Online]. <http://dermatlas.org>
- [69] An Atlas of Clinical Dermatology. [Online]. <http://dandem-pdv.is.kkh.dk/atlas>
- [70] DermNetNZ. [Online]. <http://dermnet.org.nz/lesions>
- [71] A. Jukić, I. Kopriva, and A. Cichocki, "Tensor-decomposition-based feature extraction for noninvasive diagnosis of melanoma from the clinical color image," *unpublished manuscript*, 2012.