

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Matematički odsjek

Prepoznavanje znakovnog jezika pomoću tenzora

Tim 4M

Mateo Martinjak
Mateja Pejić
Mate Poljak
Mihaela Zima

Profesor:
Zlatko Drmač

Zagreb, veljača 2022.

Sadržaj

1	Uvod	3
2	Teorija tenzora	3
3	Modeliranje znakovnog jezika pomoću tenzora	5
3.1	Trening faza	5
3.2	Test faza	7
4	Eksperiment	8

1 Uvod

Cilj ovog projekta je koristeći teoriju tenzora napraviti efektivan program za prepoznavanje znakovnog jezika. Treba imati na umu da postojeće metode prepoznavanja znakovnog jezika ovise o točki gledišta. Ova metoda ima široku primjenu, može se koristiti npr. i kod pametnih televizora gdje pokretom ruke možemo zadavati određene naredbe. Za što veću efikasnost metode bolje je prikupljati podatke iz različitih kuteva da se izbjegne ograničenje na određeni prostor i kut. Našu metodu smo primijenili na prepoznavanje nekih jednostavnih izraza znakovnog jezika.

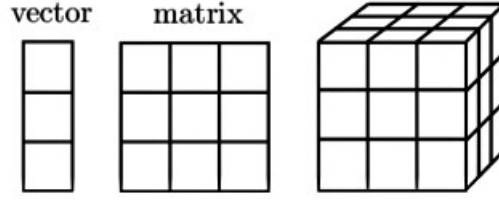


Slika 1: Abeceda

2 Teorija tenzora

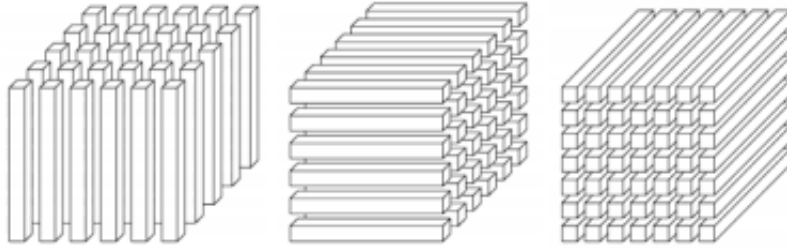
U ovom seminaru sa malim ukošenim slovima (a, b, \dots) ćemo označavati skalare, sa malim podebljanim slovima ($\mathbf{a}, \mathbf{b}, \dots$) vektore, sa velikim kaligrafskim slovima ($\mathcal{A}, \mathcal{B}, \dots$) tenzore, a sa velikim podebljanim slovima ($\mathbf{A}, \mathbf{B}, \dots$) matrice.

Definicija 1. Ako je polje \mathcal{A} određeno s N indeksa onda je $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ tenzor reda N nad \mathbb{R} , $\mathcal{A} = (x_{i_1 i_2 \dots i_N})$, $i_1 = 1, \dots, I_1, \dots, i_N = 1, \dots, I_N$.



Slika 2: Vektor je tenzor reda 1, dok je matrica tenzor reda 2. U našoj metodi radit ćemo s tenzorom reda 3, $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$.

Definicija 2. Nit (mod- n vektor) tenzora je I_n -dimenzionalni vektor dobiven fiksiranjem svakog indeksa osim indeksa I_n .



Slika 3: Niti u svim modovima tenzora reda 3.

Definicija 3. Odsječak tenzora je $2d$ -sekcija tenzora dobivena fiksiranjem svih osim dva indeksa.

Definicija 4. Mod- n produkt tenzora $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ sa matricom $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, u oznaci $\mathcal{A} \times_n \mathbf{U}$, je $(I_1 \times \dots \times I_{n-1} \times J_n \times I_{n+1} \times \dots \times I_N)$ -tenzor čije su ulazne vrijednosti dane

$$(\mathcal{A} \times_n \mathbf{U})_{i_1 i_2 \dots i_{n-1} j_n i_{n+1} \dots i_N} = \sum_{i_n} a_{i_1 i_2 \dots i_{n-1} i_n i_{n+1} \dots i_N} u_{j_n i_n} \quad (1)$$

Neka je dan tenzor $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ i matrice $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, $\mathbf{V} \in \mathbb{R}^{J_m \times I_m}$. Tada

$$(\mathcal{A} \times_n \mathbf{U}) \times_m \mathbf{V} = (\mathcal{A} \times_m \mathbf{V}) \times_n \mathbf{U} = \mathcal{A} \times_n \mathbf{U} \times_m \mathbf{V} \quad (2)$$

Slično, mod- N SVD je generalizacija SVD-a za više redove matrica. Ako je \mathcal{D} tenzor reda n i $\mathcal{D} \in \mathbb{R}^{I_1 \times \dots \times I_N}$, primjena mod- n SVD ortogonalizira pripadna polja vektora tenzora \mathcal{D} i rastavlja tenzor na sljedeći način

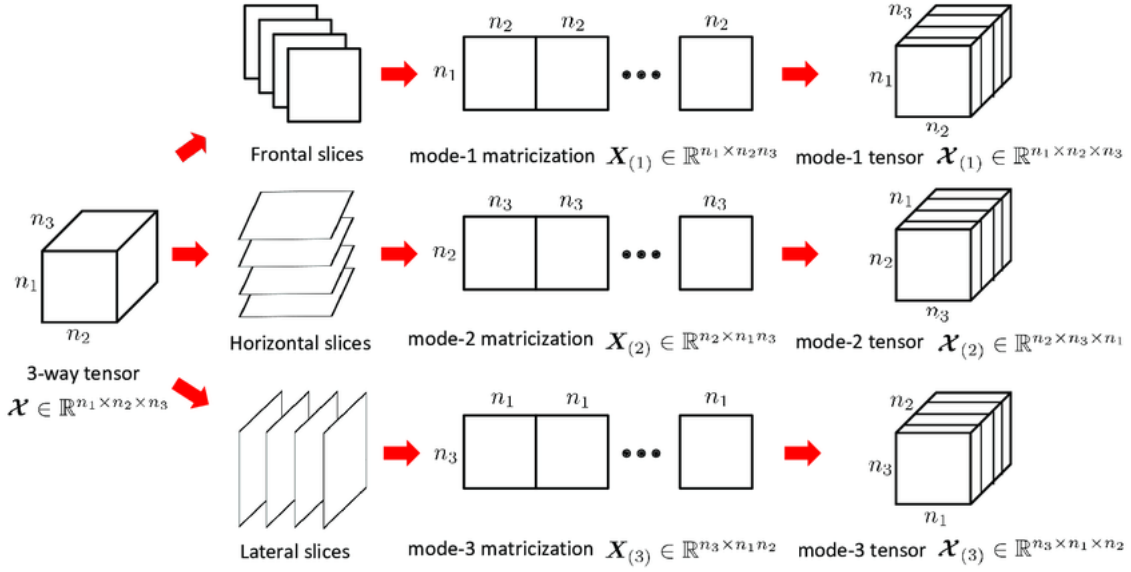
$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \dots \times_n \mathbf{U}_n \dots \times_N \mathbf{U}_N \quad (3)$$

gdje je \mathbf{U}_n , $\forall n \in \{1, 2, \dots, N\}$, ortogonalna matrica i sadrži poredane glavne komponente za mod n . \mathcal{Z} je jezgreni tenzor.

Algoritam dekompozicije provodimo u 2 koraka:

1. Za $n = 1, \dots, N$, izračunajte matricu \mathbf{U}_n u (3) računajući SVD \mathbf{D}_n , gdje je \mathbf{D}_n matricizacija tenzora u modu n , a \mathbf{U}_n je lijeva matrica SVD-a.
2. Računamo jezgreni tenzor koristeći sljedeću formulu

$$\mathcal{Z} = \mathcal{D} \times_1 \mathbf{U}_1^\top \cdots \times_n \mathbf{U}_n^\top \cdots \mathbf{U}_N^\top$$



Slika 4: Frontalni, horizontalni i bočni odječak tenzora reda 3 i pripadne matricizacije.

3 Modeliranje znakovnog jezika pomoću tenzora

3.1 Trening faza

Sve naše fotografije pohranit ćemo u jedan tenzor $\mathcal{D} \in \mathbb{R}^{I_l \times I_v \times I_{pix}}$, gdje I_l , I_v i I_{pix} označavaju broj znakova, kutova kamere i piksela slike respektivno. Pomoću HOSVD-a tenzor \mathcal{D} možemo zapisati kao

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_l \times_2 \mathbf{U}_v \times_3 \mathbf{U}_{pix} \quad (4)$$

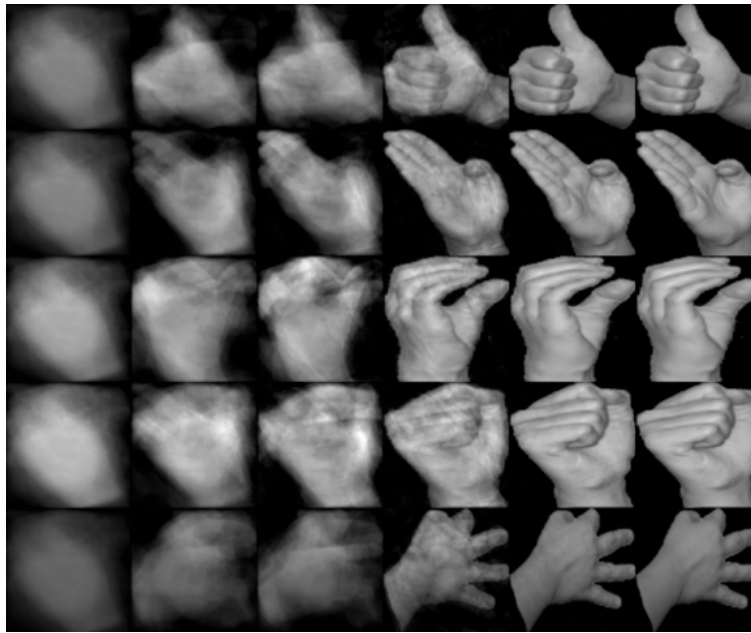
gdje je $\mathcal{Z} \in \mathbb{R}^{I_l \times I_v \times I_{pix}}$ jezgreni tenzor, dok $\mathbf{U}_l \in \mathbb{R}^{I_l \times I_l}$ i $\mathbf{U}_v \in \mathbb{R}^{I_v \times I_v}$ obuhvaćaju prostor parametara raznih znakova i kuteva kamera, respektivno. Matrica $\mathbf{U}_{pix} \in \mathbb{R}^{I_{pix} \times I_{pix}}$ predstavlja prostor svojstvenih vektora fotografija.

Red l' u matrici \mathbf{U}_l , koji označavamo sa $\mathbf{u}_l^{(l')}$, predstavlja vektor koeficijenata za znak l' , dok v' -ti redak matrice \mathbf{U}_v , koji označavamo sa $\mathbf{u}_v^{(v')}$, predstavlja vektor koeficijenata kuta kamere v' .

Pomoću formule (4) trening slika $\mathcal{D}^{(l',v')} \in \mathbb{R}^{1 \times 1 \times I_{pix}}$ za znak l' i kut kamere v' je

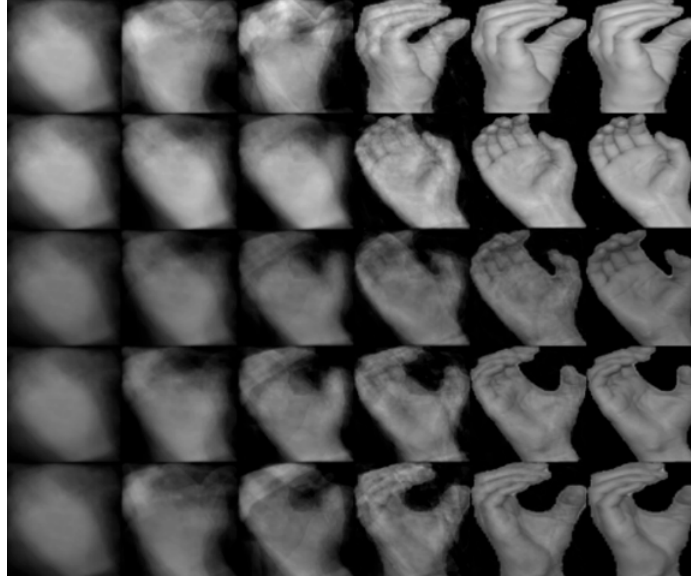
$$\mathcal{D}^{(l',v')} = \mathcal{Z} \times_1 \mathbf{u}_l^{(l')} \times_2 \mathbf{u}_v^{(v')} \times_3 \mathbf{U}_{pix}$$

Ukoliko želimo rekonstruirati fotografije iz nekog fiksnog kuta gledišta (kamere) fiksiramo v' , variramo l' , te odabiremo razne svojstvene vektore iz matrice \mathbf{U}_{pix} (recimo u rastućem poretku za što "čišću sliku").



Slika 5:

Na sličan način, ukoliko želimo prikazati fotografije za isti znak iz raznih kuteva kamere fiksiramo l' , te variramo v' i odabiremo razne svojstvene vektore iz matrice \mathbf{U}_{pix} .



Slika 6:

3.2 Test faza

Za odabranu testnu fotografiju konstruiramo tenzor \mathcal{D}_{test} dimenzije $1 \times 1 \times I_{pix}$ koji možemo zapisati kao:

$$\mathcal{D}_{test} = \mathcal{Z} \times_1 \mathbf{u}_l \times_2 \mathbf{u}_v \times_3 \mathbf{U}_{pix}$$

gdje su \mathbf{u}_l i \mathbf{u}_v prostori znakova i kuteva kamere, respektivno. Naš je cilj pronaći l' i v' koji zadovoljavaju:

$$\arg \min_{l', v'} ||\mathcal{D}_{test} - \mathcal{D}^{(l', v')}||$$

Točnije, moramo pronaći vektor iz prostora \mathbf{u}_l i \mathbf{u}_v koji će nam dati informaciju iz kojeg kuta kamere i koji znak je prisutan na testnoj fotografiji. Oni se mogu izračunati prema formuli

$$\arg \min_{\mathbf{u}_l^{(l')}, \mathbf{u}_v^{(v')}} ||\mathcal{D}_{test} - \mathcal{Z} \times_1 \mathbf{u}_l^{(l')} \times_2 \mathbf{u}_v^{(v')} \times_3 \mathbf{U}_{pix}|| \quad (5)$$

Iako nas samo zanima $\mathbf{u}_l^{(l')}$ jer nam je samo bitno koji je znak na fotografiji vektor $\mathbf{u}_v^{(v')}$ se također mora izračunati. Definirajmo sa $\{\mathbf{u}_v\}$ skup svih vektora (redova) matrice \mathbf{U}_v čija je veličina I_v . Fiksiramo li vektor $\mathbf{u}_v^{(k_v)} \in \{\mathbf{u}_v\}$ formula (5) se može napisati kao

$$\arg \min_{\mathbf{u}_l^{(l')}} ||\mathcal{D}_{test} - \mathcal{Z} \times_1 \mathbf{u}_l^{(l')} \times_2 \mathbf{u}_v^{(k_v)} \times_3 \mathbf{U}_{pix}|| \quad (6)$$

Gornja formula (6) se može napisati i kao

$$\arg \min_{\mathbf{u}_l^{(l')}} ||(D_{test}) - \mathbf{u}_l^{(l')} \times (\mathcal{Z} \times_2 \mathbf{u}_v^{(k_v)} \times_3 \mathbf{U}_{pix})||$$

Gornji minimum će se postići kada je norma jednaka 0 odnosno, za vektor $\mathbf{u}_l^{k_v}$ koji je oblika

$$\mathbf{u}_l^{k_v} = \mathcal{D}_{test} \times (\mathcal{Z} \times_2 \mathbf{u}_v^{k_v} \times_3 \mathbf{U}_{pix})^+$$

gdje eksponent "+" označava Moore-Penroseov pseudoinverz. Na ovaj način smo za svaki kut kamere dobili jedan vektor koji nam može poslužiti za određivanje znaka na fotografiji.

Nakon što dobijemo sve moguće vektore $\mathbf{u}_l^{k_v}$ koristimo kosinus formulu za računanje udaljenosti svakog vektora $\mathbf{u}_l^{k_v}$ i svakog vektora $\mathbf{u}_l^{(p)}$:

$$\arg \max_{p, k_v} \frac{\langle \mathbf{u}_l^{k_v}, \mathbf{u}_l^{(p)} \rangle}{||\mathbf{u}_l^{k_v}|| \cdot ||\mathbf{u}_l^{(p)}||}, \quad k_v = 1, \dots, I_v.$$

Dobiveni p nam govori da je na slici znak p , dok v sadrži informaciju iz kojeg kuta kamere je fotografija snimljena.

4 Eksperiment

Naš program radi za 6 znakova - "hello", "yes", "no", "thanks", "play" i "I love you". Za svaki znak smo koristili 13 fotografija za trening fazu i 2 fotografije za test fazu.



Slika 7: hello



Slika 8: love you



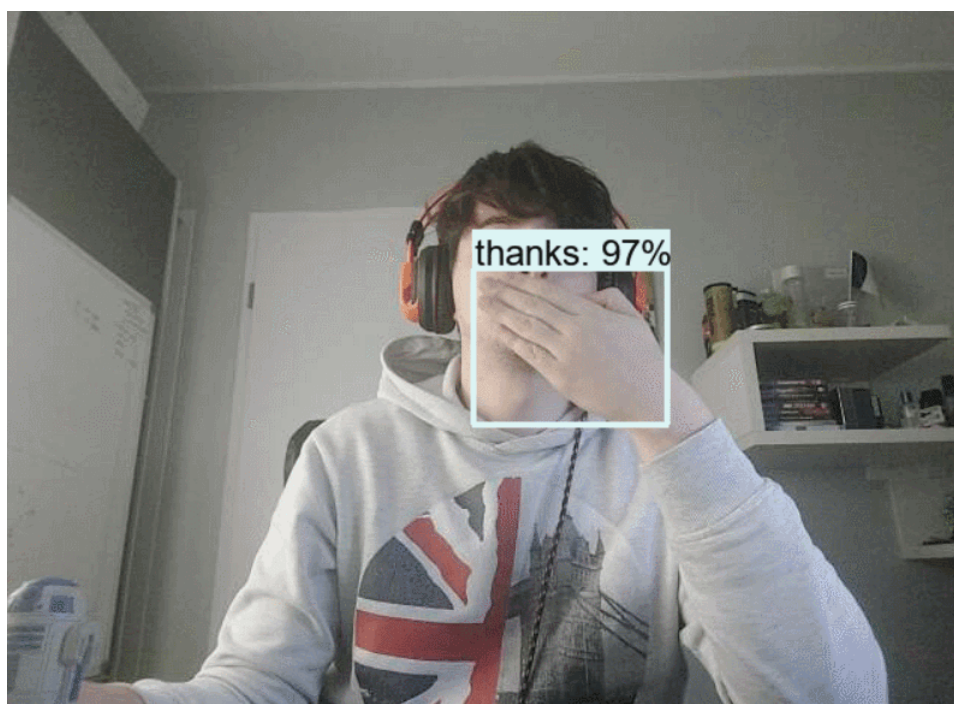
Slika 9: no



Slika 10: yes



Slika 11: play



Slika 12: thanks