Prepoznavanje rukom pisanih znamenki koristeći tenzorski SVD

Marina Matešić, Tomislav Novak, Timotej Repak

2. ožujka 2025.

Sažetak

U ovom radu prikazat ćemo 2 algoritma koji s visokom razinom uspješnosti klasificiraju rukom pisane znamenke, što je jedan od osnovnih problema klasifikacije čije primjene sežu puno dalje od same klasifikacije znamenki. Oba algoritma koriste tenzorske metode, posebice HOSVD.

Sadržaj

1 Osnovno o tenzorima i potrebni rezultati

U algoritmima koji se koriste u ovom radu, bit će potrebni samo tenzori reda 3 pa ćemo sve općenitije rezultate prikazati u specijalnom slučaju za takve tenzore. U ostatku ovog rada tenzore reda 3 nazivat ćemo samo tenzorima.

Tenzor možemo shvatiti kao trodimenzionalno polje realnih brojeva, tj. tenzor $\mathcal{A} = (a_{ijk}) \in \mathbb{R}^{I\times J\times K}$, gdje su I, J i K prirodni brojevi, a dimenzija vektorskog prostora $\mathbb{R}^{I\times J\times K}$ je IJK.

Za dva tenzora $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I\times J\times K}$ definiramo njihov skalarni produkt kao

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} a_{ijk} b_{ijk}. \tag{1}$$

Kažemo da su $\mathcal{A}$ i $\mathcal{B}$ ortogonalni ako vrijedi $\langle \mathcal{A}, \mathcal{B} \rangle = 0$.

Analogno vektorskom slučaju, normu tenzora $\mathcal{A}$ definiramo kao $\|\mathcal{A}\| = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}$.

Dijelove tenzora $\mathcal{A}(:, j, k)$, $\mathcal{A}(i, :, k)$ i $\mathcal{A}(i, j, :)$ nazivamo niti u modu 1, 2 i 3, redom.

Matricizacija u n-tom modu tenzora $\mathcal{A}$ je matrica $A_{(n)}$ koju dobijemo tako da niti od $\mathcal{A}$ stavimo kao retke matrice $A_{(n)}$.

n-rang tenzora $\mathcal{A}$ definiramo kao $\operatorname{rank}_n(\mathcal{A}) = \operatorname{rank}(A_{(n)})$. Lako se uoči da općenito ne vrijedi $\operatorname{rank}_1(\mathcal{A}) = \operatorname{rank}_2(\mathcal{A}) = \operatorname{rank}_3(\mathcal{A})$.

Množenjem tenzora $\mathcal{A} \in \mathbb{R}^{I\times J\times K}$ matricom $F \in \mathbb{R}^{L\times I}$ u modu 1 dobije se tenzor $\mathbb{R}^{L\times J\times K} \ni \mathcal{B} = \mathcal{A} \times_1 F, \quad \mathcal{B}(l, j, k) = \sum_{i=1}^{I} \mathcal{A}(i, j, k) F(l, i)$.

Množenjem tenzora $\mathcal{A} \in \mathbb{R}^{I\times J\times K}$ matricom $F \in \mathbb{R}^{L\times J}$ u modu 2 dobije se tenzor $\mathbb{R}^{I\times L\times K} \ni \mathcal{B} = \mathcal{A} \times_2 F, \quad \mathcal{B}(i, l, k) = \sum_{j=1}^{J} \mathcal{A}(i, j, k) F(l, j)$.

Množenjem tenzora $\mathcal{A} \in \mathbb{R}^{I\times J\times K}$ matricom $F \in \mathbb{R}^{L\times K}$ u modu 3 dobije se tenzor $\mathbb{R}^{I\times J\times L} \ni \mathcal{B} = \mathcal{A} \times_3 F, \quad \mathcal{B}(i, j, l) = \sum_{k=1}^{K} \mathcal{A}(i, j, k) F(l, k)$.

Za tenzor $\mathcal{A} \in \mathbb{R}^{I\times J\times K}$ i matrice $F \in \mathbb{R}^{L\times I}$, $G \in \mathbb{R}^{M\times J}$ i $H \in \mathbb{R}^{N\times IL}$ vrijedi $(\mathcal{A} \times_1 F) \times_2 G = (\mathcal{A} \times_2 G) \times_1 F = \mathcal{A} \times_1 F \times_2 G \in \mathbb{R}^{L\times M\times K}$ te $(\mathcal{A} \times_1 F) \times_1 H = \mathcal{A} \times_1 (HF) \in \mathbb{R}^{N\times J\times K}$.

Tenzorski SVD - HOSVD: Tenzor $\mathcal{A} \in \mathbb{R}^{I\times J\times K}$ se može zapisati kao produkt

$$\mathcal{A} = \mathscr{S} \times_1 U \times_2 V \times_3 W, \tag{2}$$

gdje su:

(1) $U \in \mathbb{R}^{I\times I}$, $V \in \mathbb{R}^{J\times J}$ i $W \in \mathbb{R}^{K\times K}$ ortogonalne matrice,

(2) $\mathscr{S}$ je realni tenzor istih dimenzija kao i $\mathcal{A}$ sa sljedećim svojstvima:

   (a) (potpuna ortogonalnost) svaka dva odsječka istog tipa su okomiti, tj.

$$\langle \mathscr{S}(v,:,:), \mathscr{S}(\lambda,:,:)\rangle = 0 \quad \text{za } v \neq \lambda,$$
$$\langle \mathscr{S}(:,v,:), \mathscr{S}(:,\lambda,:)\rangle = 0 \quad \text{za } v \neq \lambda,$$
$$\langle \mathscr{S}(:,:,v), \mathscr{S}(:,:,\lambda)\rangle = 0 \quad \text{za } v \neq \lambda.$$

   (b) (*uređenost*) norme odsječaka u svakom modu su uređene (poredane), npr. za mod 1 vrijedi $\|\mathscr{S}(1,:,:)\| \geq \|\mathscr{S}(2,:,:)\| \geq \cdots \geq 0$. Te norme u modu n su upravo singularne vrijednosti matriciziranog tenzora $A_{(n)}$.

Zbog svojstva uređenosti, tenzor koji je prikazan u HOSVD formatu aproksimiramo tenzorom manjih dimenzija tako da ga prikažemo kao produkt gore navedenog tenzora i matrica, ali sa "odsječenim" dijelovima koji su dalje od pozicije $(1,1,1)$, odnosno $(1,1)$.

Svaki se tenzor može zapisati u obliku $\mathcal{A} = \sum_{v=1}^{K} A_v \times_3 w_v$, gdje je $A_v = \mathcal{S}(:,:,v) \times_1 U \times_2 V$ te vrijedi $\langle A_v, A_\mu \rangle = 0$.

2 Algoritmi

2.1 Klasifikacija pomoću HOSVD-a

Neka su u bazi rukom pisanih znamenaka sve znamenke prikazane kao matrice tipa $I \times J$ te neka svakih znamenki ima K (ne nužno, ali za trenutnu jednostavnost zapisa). Sve matrice koje predstavljaju istu znamenku onda posložimo u tenzor čiji su one frontalni slice-ovi.

Na primjer, tenzor koji predstavlja broj 7 neka je $\mathcal{A}^7 \in \mathbb{R}^{I\times J\times K}$ te neka imamo njegovu HOSVD dekompoziciju. Po posljednjoj napomeni iz prethodnog poglavlja, taj tenzor zapišemo kao $\mathcal{A}^7 = \sum_{v=1}^{K} A_v^7 \times_3 w_v^7$ što znači da je svaka sedmica na jedinstven način prikazana kao kombinacija matrica A_v^7 koje čine bazu.

Za neki k uzmimo dominantni k-dimenzionalni potprostor za svaki tenzor koji predstavlja znamenku i označimo $T^\mu = (A_v^\mu)_{v=1}^k, \quad \mu = 0, 1, \ldots, 9$. To možemo zato što su norme od matrica te baze redom padajuće, pa ćemo dobiti na vremenu/složenosti uzimajući manju bazu (potprostor manje dimenzije) kao aproksimaciju a pritom odbacujući 'najnebitniji' dio (onaj s najmanje informacija).

Neka je D neka znamenka koju treba klasificirati, normalizirana. Naš je cilj odrediti T^μ koji najbolje opisuje D pa ćemo onda klasificirati D kao matricu koja odgovara znamenki μ.

Pod "najbolje opisuje" smatrat ćemo da najbolje odgovara u smislu najmanjih kvadrata, odn. vidjet ćemo kojim potprostorom možemo opisati matricu D s najmanjom greškom.

Dakle, rješavamo minimizacijsku zadaću $\min_{\alpha_v^\mu} \|D - \sum_{v=1}^k \alpha_v^\mu A_v^\mu\|$ čije je rješenje zbog ortonormiranosti dano s $\hat{\alpha}_v^\mu = \langle D, A_v^\mu \rangle$. Odnosno, tražimo μ koji minimizira funkciju $R(\mu) = \|D - \sum_{v=1}^k \hat{\alpha}_v^\mu A_v^\mu\|^2 = \left\langle D - \sum_{v=1}^k \hat{\alpha}_v^\mu A_v^\mu, D - \sum_{v=1}^k \hat{\alpha}_v^\mu A_v^\mu \right\rangle = \langle D, D\rangle - \sum_{v=1}^k \langle D, A_v^\mu\rangle^2 = 1 - \sum_{v=1}^k \langle D, A_v^\mu\rangle^2$.

2.2 Klasifikacija pomoću HOSVD-a s kompresijom

Kako bismo smanjili složenost, u ovom algoritmu koristit ćemo kompresiju podataka prije izgradnje modela klasa.

2.2.1 Trening faza

Neka je $\mathcal{D}$ tenzor koji sadrži sve znamenke iz trening skupa. Budući da su znamenke grupirane po klasama, možemo zapisati HOSVD dekompoziciju cijelog tenzora:

$$\mathcal{D} = \mathscr{S} \times_1 U \times_2 V \times_3 W \approx \mathcal{F} \times_1 U_p \times_2 V_q, \tag{3}$$

gdje je $U_p = U(:, 1:p)$, $V_q = V(:, 1:q)$ i $\mathcal{F} = \mathscr{S}(1:p, 1:q, :) \times_3 W$.

Ovom aproksimacijom smanjujemo dimenzionalnost prikaza znamenaka na $\mathbb{R}^p$ te broj znamenaka u svakoj klasi na q. Reducirani tenzor $\mathcal{F}$ ima oblik:

$$\mathcal{F} = \mathcal{D} \times_1 U_p^T \times_2 V_q^T. \tag{4}$$

Uz pretpostavku da su p i q znatno manji od dimenzija tenzora $\mathcal{D}$, postiže se značajno smanjenje skupa podataka.

Niskodimenzionalna reprezentacija znamenaka dobiva se kao:

$$D_p = D \times_1 U_p^T = \mathcal{F} \times_2 V_q, \tag{5}$$

gdje su stupci D_p reprezentacije znamenaka, a odsječci $\mathcal{F}$ sadrže bazne vektore za svaku klasu.

Nakon kompresije, za svaku znamenku μ formiramo matricu $F^\mu := \mathcal{F}(:, :, \mu)$ koja sadrži samo znamenke klase μ. SVD dekompozicijom dobivamo ortonormirane baze B^μ za svaku klasu:

$$F^\mu = [B^\mu (B^\mu)^\perp] \Sigma^\mu (Q^\mu)^T, \quad \mu = 0, 1, \ldots, 9, \tag{6}$$

2.2.2 Faza testiranja

Za testiranje, ulaznu znamenku D projiciramo u reducirani prostor pomoću U_p:

$$D_p = U_p^T D. \tag{7}$$

Nakon toga rješavamo niz problema najmanjih kvadrata za svaku znamenku μ:

$$\min_{x^\mu} \|D_p - B^\mu x^\mu\|. \tag{8}$$

Budući da su stupci B^μ ortonormalni, rješenje je dano s:

$$\hat{x}^\mu = (B^\mu)^T D_p. \tag{9}$$

Konačno, znamenku svrstamo u klasu μ za koju je rezidual najmanji:

$$R(\mu) = \|D_p - B^\mu (B^\mu)^T D_p\|. \tag{10}$$

3 Tehnička izvedba algoritma i analiza složenosti

3.1 Tehnička izvedba

Algoritam implementiramo u programskom jeziku Python te isječke koda ne prilažemo u sami rad jer su vrlo interaktivno

4 Rezultati i zaključak