UCLA

Neural Nonnegative CP Decomposition for Hierarchical Tensor Analysis

Joshua Vendrow, Jamie Haddock, Deanna Needell

Department of Mathematics, UCLA

Background

- **Nonnegative Matrix Factorization (NMF)** is an approach typically applied in unsupervised tasks such as dimensionality-reduction, latent topic modeling, and clustering.
- Given nonnegative data matrix $X \in_{\geq 0}^{m \times n}$ and a user-defined target dimension $r \in \mathbb{N}$, NMF seeks nonnegative factor matrices $A \in_{\geq 0}^{m \times r}$, and $S \in_{\geq 0}^{r \times n}$ such that $X \approx AS$, formulated as

$$\underset{A \geq 0, S \geq 0}{\arg\min} \|X - AS\|_F^2. \tag{1}$$

- **Nonnegative CP Decomposition (NCPD)** generalizes NMF to multi-model tensor data.
- Given nonnegative data tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times \ldots \times n_k}$ and a user-defined target dimension $r \in \mathbb{N}$, NCPD seeks nonnegative factor matrices $X_1, X_2, \ldots, X_k$ where $X_i \in_{\geq 0}^{n_i \times r}$ such that $\mathcal{X} \approx [[X_1, X_2, \cdots, X_k]] \equiv \sum_{j=1}^{r} x_j^{(1)} \otimes x_j^{(2)} \otimes \cdots \otimes x_j^{(k)}$, where $x_j^{(i)}$ is the jth column of X_i.

Method

- Given a nonnegative order-k tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times \ldots \times n_k}$, **Hierarchical NCPD (HNCPD)** consists of an initial rank-r NCPD layer with factor matrices $X_1, X_2, \ldots, X_k$, each with r columns, and an HNMF with ranks $r^{(0)}, r^{(1)}, \cdots, r^{(\mathcal{L}-2)}$ for each of these factors matrices.
- For each X_i at layer ℓ, we factorize X_i as

$$X_i \approx \widetilde{X}_i \equiv A_i^{(0)} A_i^{(1)} \ldots A_i^{(\ell-2)} S_i^{(\ell-2)}. \tag{2}$$

- We propose **Neural NCPD**, a method for training an HNCPD model by representing the model with a neural network architecture.
- Our iterative method consists of two subroutines, a forward-propagation and a backpropagation. In Algorithms 1 and 2, we display the pseudocode for our proposed method.

Algorithm 1 Forward Propagation

procedure FORWARDPROP($\{X_i\}_{i=0}^k$, $\{A_i^{(\ell)}\}_{i=0,\ell=0}^{k,\mathcal{L}-2}$)
  for $i = 1, \cdots, k$ **do**
    for $\ell = 0, \cdots, \mathcal{L} - 2$ **do**
      $S_i^{(\ell)} \leftarrow \arg\min_{S \geq 0} \|S_i^{(\ell-1)} - A_i^{(\ell)} S\|_F$
    end for
  end for
end procedure

Algorithm 2 Neural NCPD

Input: Tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_k}$, cost C
$X_1, X_2, \ldots, X_k \leftarrow$ NCPD($\mathcal{X}$), initialize $\{A_i^{(\ell)}\}_{i=0,\ell=0}^{k,\mathcal{L}-2}$
for iterations $= 1, \ldots, T$ **do**
  ForwardProp($\{X_i\}_{i=0}^k$, $\{A_i^{(\ell)}\}_{i=0,\ell=0}^{k,\mathcal{L}-2}$)
  for $i = 1, \cdots, k$, $\ell = 0, \cdots, \mathcal{L} - 2$ **do**
    $A_i^{(\ell)} \leftarrow \left(\texttt{optimizer}\left(A_i^{(\ell)}, \frac{\partial C}{\partial A_i^{(\ell)}}\right)\right)^+$
  end for
end for

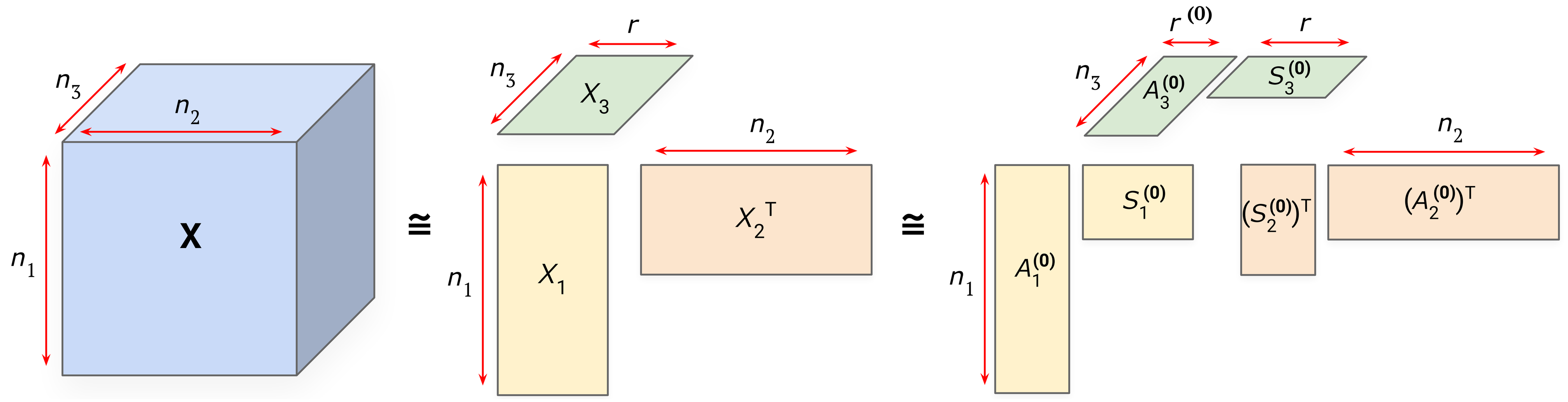

Figure 1: Visualization of a two-layer HNCPD model.

Conclusions

In this paper, we introduced the hierarchical NCPD model and presented a novel method, Neural NCPD, to train this decomposition. We empirically demonstrate the promise of this method on both real and synthetic datasets; in particular, this model reveals the hierarchy of topics learned at different NCPD ranks, which is not available to standard NCPD or NMF-based approaches.

Contact Information

jvendrow@math.ucla.edu
www.joshvendrow.com
github.com/jvendrow

Results

- We generate a synthetic tensor (see Figure 2) with overlapping and non-overlapping blocks of varying size and intensity to form a hierarchical structure.
- We see in Figure 3 that Neural NCPD is better able to identify the topic structure of the underlying model.

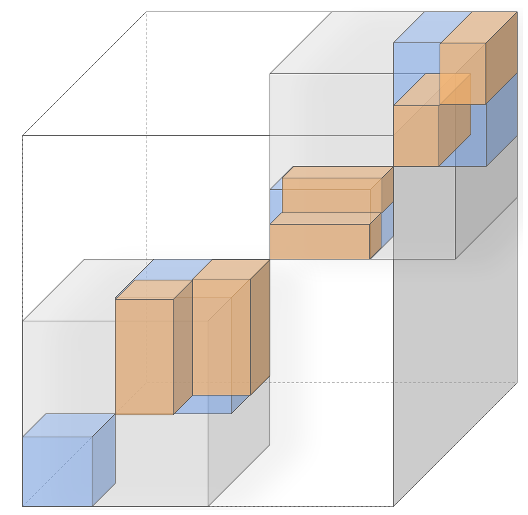

Figure 2: Visualization of the synthetic tensor.

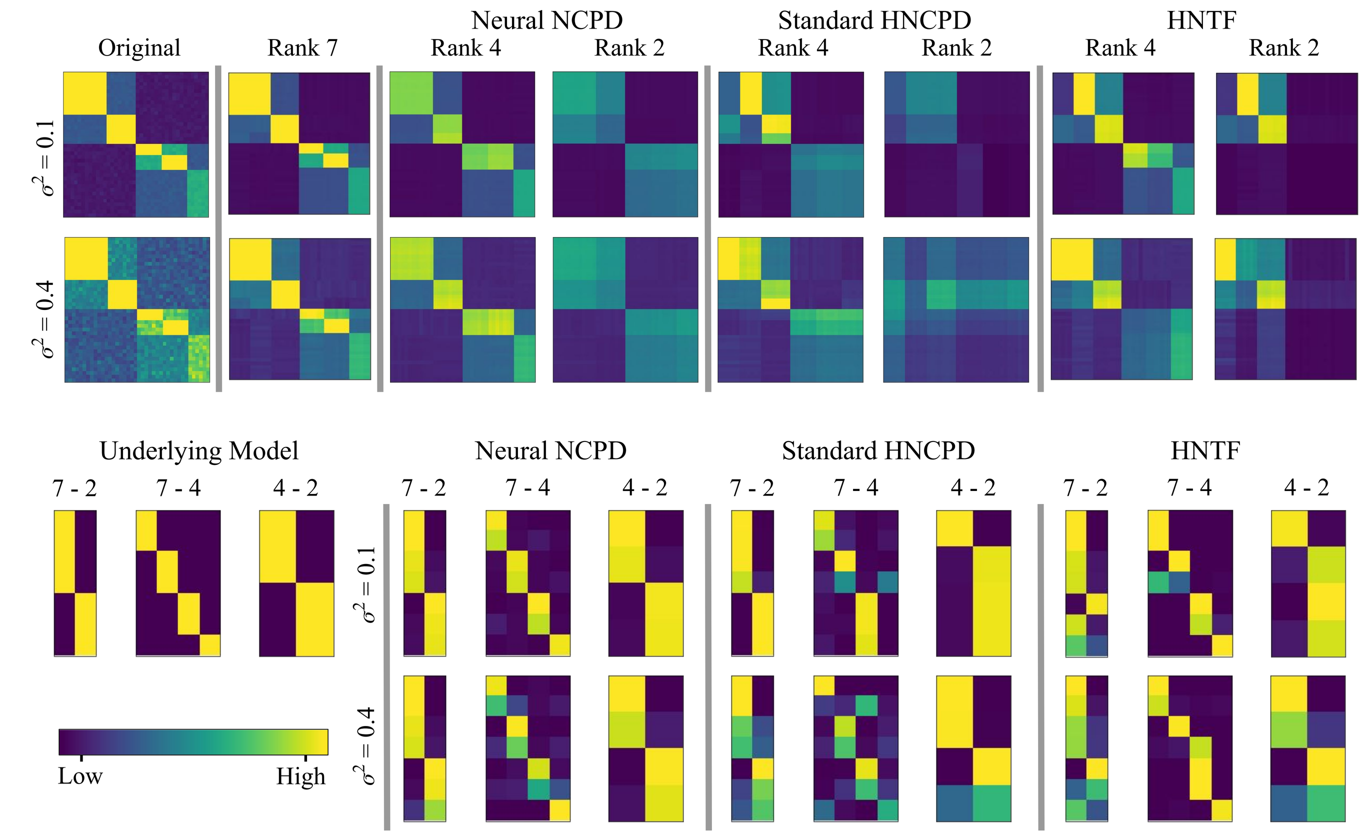

Figure 3: (Top left) Data tensor $\mathcal{X}$ with two levels of noise. (Top right) ranks 7, 5, and 3 Neural NCPD, Standard HNCPD, and HNTF approximations of $\mathcal{X}$. (Bottom left) Underlying topic modelling matrix. (Bottom right) topic modelling matrices for each method.

Table 1: Topic modeling loss and relative reconstruction on the synthetic dataset by NCPD and NMF methods.

		Topic Modeling Loss						Relative Reconstruction Loss					
		$\sigma^2 = 0.1$			$\sigma^2 = 0.4$			$\sigma^2 = 0.1$			$\sigma^2 = 0.4$		
Method	Mode	$7-2$	$7-4$	$4-2$	$7-2$	$7-4$	$4-2$	$r=7$	$r^{(0)}=4$	$r^{(1)}=2$	$r=7$	$r^{(0)}=4$	$r^{(1)}=2$
Neural HNCPD		**0.043**	0.042	**0.042**	**0.087**	**0.087**	**0.081**	0.119	0.252	0.563	**0.454**	0.508	0.714
Standard HNCPD		0.106	0.101	0.189	0.145	0.193	0.204	0.119	0.494	0.828	**0.454**	0.612	0.892
HNTF	1	0.163	0.236	0.182	0.171	0.144	0.170	0.119	0.502	0.795	**0.454**	0.576	0.781
	2	0.087	**0.040**	0.101	0.090	0.116	0.142	0.119	0.309	0.665	**0.454**	0.587	0.765
	3	0.078	0.122	0.106	0.084	0.111	0.164	0.119	0.417	0.713	**0.454**	0.560	0.747
Neural NMF	1	0.154	0.192	0.105	0.169	0.219	0.127	0.146	0.268	0.593	0.478	0.521	0.705
	2	0.075	0.244	0.146	0.153	0.190	0.160	0.141	0.289	0.585	0.475	0.513	0.710
	3	0.119	0.164	0.110	0.158	0.197	0.140	0.151	0.236	0.576	0.477	0.512	**0.693**
Standard HNMF	1	0.098	0.182	0.052	0.164	0.219	0.139	**0.118**	0.235	**0.558**	0.472	0.524	0.707
	2	0.080	0.199	0.090	0.151	0.213	0.088	**0.118**	0.245	0.566	0.472	0.505	0.709
	3	0.060	0.165	0.085	0.137	0.193	0.114	**0.118**	**0.233**	0.563	0.472	**0.503**	0.717

- From Table 1, we see that the the topic modelling loss for Neural NCPD is less than that of every other NCPD and NMF model in all but one case.
- Reconstruction loss for Neural NCPD is significantly better than that of Standard HNCPD and HNTF across rank and level of noise.

- The **Twitter political data set** is a data set of tweets sent by eight political candidates during the 2016 presidential election season.
- We bin tweets made by a candidate within each month over 10 months, resulting in a tensor of size $8 \times 10 \times 12721$.
- In Figure 4 we visualize topics learned by Neural NCPD and see meaningful hierarchical relationships; e.g., Cruz and Kasich, who left the race at similar times, are grouped together than rank 4.

Table 2: Relative reconstruction loss on the Twitter political dataset.

Method	$r = 8$	$r^{(0)} = 4$	$r^{(1)} = 2$
Neural NCPD	0.834	**0.883**	**0.918**
Standard NCPD	0.834	0.889	0.919
Standard HNCPD	0.834	0.913	0.976
HNTF-1	0.834	0.890	0.927
HNTF-2	0.834	0.909	0.956
HNTF-3	0.834	0.895	0.942

Rank 8 Topics

Topic 1	Topic 2	Topic 3	Topic 4
trump	senate	martinomalley	berniesanders
hillary	florida	hillaryclinton	people
donald	zika	realdonaldtrump	bernie
president	venezuela	campaigning	must
timkaine	nicolasmaduro	maryland	change

Topic 5	Topic 6	Topic 7	Topic 8
tedcruz	johnkasich	marcorubio	crooked
cruz	kasich	teammarco	hillary
ted	ohio	vote	thank
internet	john	flsen	great
choosecruz	gov	click	clinton

Rank 4 Topics

Topic 1	Topic 2
marcorubio	trump
teammarco	hillary
vote	donald
flsen	people
click	vote

Topic 3	Topic 4
tedcruz	senate
cruz	florida
ted	zika
johnkasich	venezuela
kasich	nicolasmaduro

Rank 2 Topics

Topic 1
trump
hillary
vote
people
donald

Topic 2
tedcruz
cruz
ted
johnkasich
kasich

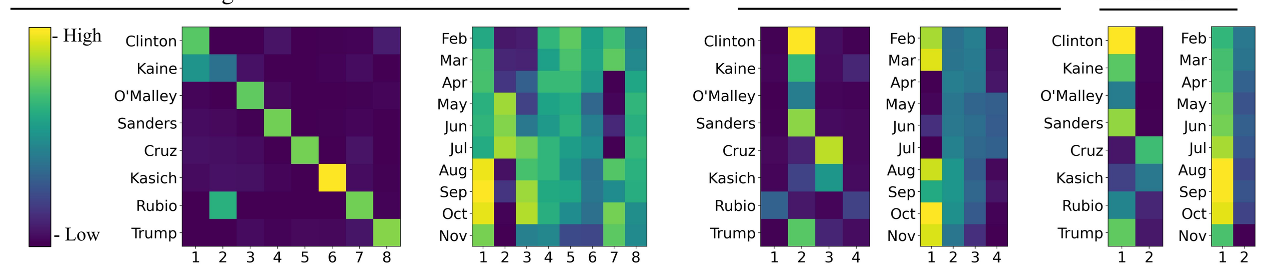

Figure 4: A three-layer Neural NCPD on the Twitter dataset at ranks $r = 8$, $r^{(0)} = 4$ and $r^{(1)} = 2$. At each rank, we display the top keywords and topic heatmaps for candidate and temporal modes.

Acknowledgements

The authors are grateful to and were partially supported by NSF CAREER DMS #1348721, NSF DMS #2011140 and NSF BIGDATA DMS #1740325. JH is also partially supported by NSF DMS #2111440.