Hierarchical and Neural Nonnegative Tensor Decompositions

by Jamie Haddock
(Harvey Mudd College, Department of Mathematics)
on December 2, 2022,
IPAM “Multi-Modal Imaging with Deep Learning and Modeling”


joint with Joshua Vendrow*, Deanna Needell

joint with Joshua Vendrow*, Deanna Needell

NSF DMS #2211318
Motivation
Learn trends in high-dimensional data

Motivation

Introduction

Hierarchical Models

Experiments

Backpropagation

Conclusions

Patient Surveys

Term-Document Matrix

... my migraines. Of course I have heart issues too, but the migraines are my main concern right now. My priority is getting that pain under control. I’m happy that my doctors are managing the pain and dizziness automatically.

... My doctor was great, realized it was a heart attack really quick. I didn’t quite know what to expect when I told the doctor about the weakness and pain. I expected them to prescribe painkillers or something like that but they didn’t prescribe anything. The pain persisted. I was still in incredible pain by the next day. I don’t know why the pain lasted so long. I just stress, but my mom had migraines. I told her about what I was feeling and she realized it was exactly what she had. Sometimes the pain is because of my heart, sometimes it’s because of my vision. The pain aggravated my high blood pressure.

... chest pain. I had been feeling lightheaded and nauseous. The pain was definitely there but really I felt more a tightness in my chest than anything. It left me short of breath, which was probably making me lightheaded. The EKG indicated that my heart had several blockages that would need a stent. My cardiologists were able to clear the blockages and I spent one night under watch in the hospital.

After my heart attack, I completely changed my lifestyle. I quit smoking, started an exercise regimen and diet...
Learn trends in high-dimensional data

Patient Surveys

Term-Document Matrix

Understand symptom trends and shared patient experiences automatically.
» Learn trends in high-dimensional data
» Learn trends in high-dimensional data
Learn trends in high-dimensional data

Learn cohesive parts and separate noise in medical image studies.
Can we tell how the resulting parts/topics are related?
Can we tell how the resulting parts/topics are related?

How do we choose the number of topics or parts to learn?
Can we tell how the resulting parts/topics are related?

How do we choose the number of topics or parts to learn?

Hierarchical matrix factorization and tensor decomposition topic models!
Introduction
Nonnegative Matrix Factorization (NMF)

**Model:** Given nonnegative data $X$, compute nonnegative $A$ and $S$ of lower rank so that

$$X \approx AS.$$
Nonnegative Matrix Factorization (NMF)

**Model:** Given nonnegative data $X$, compute nonnegative $A$ and $S$ of lower rank so that

$$X \approx AS.$$
Nonnegative Matrix Factorization (NMF)

Employed for dimensionality-reduction and topic modeling

» Nonnegative Matrix Factorization (NMF)

Employed for dimensionality-reduction and topic modeling

Often formulated as

\[
\min_{\mathbf{A}\in\mathbb{R}_{\geq 0}^{n_1\times n_2}, \mathbf{S}\in\mathbb{R}_{\geq 0}^{r\times n_2}} \|\mathbf{X} - \mathbf{AS}\|_F^2 \quad \text{or} \quad \min_{\mathbf{A}\in\mathbb{R}_{\geq 0}^{n_1\times r}, \mathbf{S}\in\mathbb{R}_{\geq 0}^{r\times n_2}} D(\mathbf{X}||\mathbf{AS}).
\]

» Nonnegative Matrix Factorization (NMF)

- Employed for dimensionality-reduction and topic modeling
- Often formulated as

\[
\min_{A \in \mathbb{R}_{\geq 0}^{n_1 \times r}, S \in \mathbb{R}_{\geq 0}^{r \times n_2}} \|X - AS\|_F^2 \quad \text{or} \quad \min_{A \in \mathbb{R}_{\geq 0}^{n_1 \times r}, S \in \mathbb{R}_{\geq 0}^{r \times n_2}} D(X \| AS)^.1
\]

- non-convex optimization problems

---

Nonnegative Matrix Factorization (NMF)

Employed for dimensionality-reduction and topic modeling.

Often formulated as

\[ \min_{A \in \mathbb{R}^{n_1 \times r} \geq 0, S \in \mathbb{R}^{r \times n_2} \geq 0} \| X - AS \|_F \]

or

\[ \min_{A \in \mathbb{R}^{n_1 \times r} \geq 0, S \in \mathbb{R}^{r \times n_2} \geq 0} D(X \parallel AS) \]

Non-convex optimization problems

Nonnegative CANDECOMP/PARAFAC (CP) decomposition (NCPD)

\[
X \approx X_1 + \cdots + X_r
\]

Nonnegative CANDECOMP/PARAFAC (CP) decomposition (NCPD)

- Formulated as \( \min_{X_i \geq 0} \| X - [X_1, X_2, \ldots, X_k] \|_F^2 \) where

\[
[X_1, X_2, \ldots, X_k] \equiv \sum_{j=1}^{r} x_j^{(1)} \otimes x_j^{(2)} \otimes \cdots \otimes x_j^{(k)}
\]

---


Hierarchical NMF

Model: Sequentially factorize

Hierarchical NMF

Model: Sequentially factorize

\[ X \approx A^{(0)} S^{(0)} \]
Hierarchical NMF

Model: Sequentially factorize

\[ X \approx A^{(0)} S^{(0)}, S^{(0)} \approx A^{(1)} S^{(1)} \]

» **Hierarchical NMF**

**Model:** Sequentially factorize

\[ X \approx A^{(0)} S^{(0)}, \quad S^{(0)} \approx A^{(1)} S^{(1)}, \quad S^{(1)} \approx A^{(2)} S^{(2)}, \ldots, \quad S^{(L-1)} \approx A^{(L)} S^{(L)}. \]

\[ \triangleright \quad k^{(\ell)} : \text{supertopics collecting } k^{(\ell-1)} \text{ subtopics} \]

---

Hierarchical NMF

Model: Sequentially factorize

\[ X \approx A^{(0)} S^{(0)}, \quad S^{(0)} \approx A^{(1)} S^{(1)}, \quad S^{(1)} \approx A^{(2)} S^{(2)}, \ldots, \quad S^{(L-1)} \approx A^{(L)} S^{(L)}. \]

\[ \Delta k^{(\ell)}: \text{supertopics collecting } k^{(\ell-1)} \text{ subtopics} \]

\[ \text{provides relationship between data matrix modes and } k^{(\ell)} \text{ topics} \]

Hierarchical NMF

Motivation

Introduction

Hierarchical Models

Experiments

Backpropagation

Conclusions
Hierarchical NMF

eulicits the hierarchical relationships of learned topics.

no need to choose a fixed model rank (number of topics).
Hierarchical NMF

Motivation
Introduction
Hierarchical Models
Experiments
Backpropagation
Conclusions
Hierarchical NMF elucidates the hierarchical relationships of learned topics.
Hierarchical NMF

- elucidates the hierarchical relationships of learned topics
- no need to choose a fixed model rank (number of topics)
<table>
<thead>
<tr>
<th>Motivation</th>
<th>Introduction</th>
<th>Hierarchical Models</th>
<th>Experiments</th>
<th>Backpropagation</th>
<th>Conclusions</th>
</tr>
</thead>
</table>

**Hierarchical Models**
Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?
Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?

Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?

Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?


Results depend upon hyperparameter choice (mode).
Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?


Results depend upon hyperparameter choice (mode).

Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?


  Results depend upon hyperparameter choice (mode).


  Not a single hierarchical relationship, good training method.
Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?


Results depend upon hyperparameter choice (mode).


Not a single hierarchical relationship, good training method.
Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?


Results depend upon hyperparameter choice (mode).

  Not a single hierarchical relationship, good training method.
  Single hierarchical relationship, naive training method.
Hierarchical NCPD Model (Take 1)

Hierarchical NCPD Model (Take 1)

Learn an initial rank-$r$ NCPD model,

$$\mathbf{X} \approx [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k]$$

Hierarchical NCPD Model (Take 1)

Learn an initial rank-$r$ NCPD model,

$$\mathbf{X} \approx [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_k]$$

and apply a hierarchical NMF model independently to each factor matrix,

$$\mathbf{X}_i \approx \mathbf{A}_i^{(0)} \mathbf{A}_i^{(1)} \cdots \mathbf{A}_i^{(l)} \mathbf{S}_i^{(l)}.$$

---

» Hierarchical NCPD Model (Take 1)

Hierarchical NCPD Model (Take 1)

- can extend good training method for HNMF (Neural NMF) → Neural NCPD (later in this talk!)

---

» Hierarchical NCPD Model (Take 1)

▷ can extend good training method for HNMF (Neural NMF) → Neural NCPD (later in this talk!)

▷ Different hierarchy across tensor modes. :( 

---

Multi-HNTF Model (Take 2)

This model learns

\[ X \approx [X^{(0)}_1, X^{(0)}_2, \ldots, X^{(0)}_k] \]
» **Multi-HNTF Model (Take 2)**

This model learns

\[
X \approx \begin{bmatrix} X_1^{(0)}, X_2^{(0)}, \ldots, X_k^{(0)} \end{bmatrix} \approx \begin{bmatrix} X_1^{(1)}, X_2^{(1)}, \ldots, X_k^{(1)} \end{bmatrix}
\]

where

\[
X_j^{(\ell+1)} = X_j^{(\ell)} W^{(\ell)} ,
\]

and \( W^{(\ell)} \in \mathbb{R}_{\geq 0}^{r^{(\ell-1)} \times r^{(\ell)}} \).

---

» **Multi-HNTF Model (Take 2)**

This model learns

\[
\mathbf{X} \approx \begin{bmatrix} \mathbf{X}_1^{(0)}, \mathbf{X}_2^{(0)}, \ldots, \mathbf{X}_k^{(0)} \end{bmatrix} \approx \begin{bmatrix} \mathbf{X}_1^{(1)}, \mathbf{X}_2^{(1)}, \ldots, \mathbf{X}_k^{(1)} \end{bmatrix}
\]

where

\[
\mathbf{X}_j^{(\ell+1)} = \mathbf{X}_j^{(\ell)} \mathbf{W}^{(\ell)},
\]

and \(\mathbf{W}^{(\ell)} \in \mathbb{R}_{\geq 0}^{r^{(\ell-1)} \times r^{(\ell)}}\).
Multi-HNTF Model (Take 2)

This model learns

\[ \mathbf{X} \approx [\mathbf{X}_1^{(0)}, \mathbf{X}_2^{(0)}, \ldots, \mathbf{X}_k^{(0)}] \approx [\mathbf{X}_1^{(1)}, \mathbf{X}_2^{(1)}, \ldots, \mathbf{X}_k^{(1)}] \approx \ldots \approx [\mathbf{X}_1^{(L-1)}, \mathbf{X}_2^{(L-1)}, \ldots, \mathbf{X}_k^{(L-1)}] \]

where

\[ X_j^{(\ell+1)} = X_j^{(\ell)} W^{(\ell)}, \]

and \( W^{(\ell)} \in \mathbb{R}_{\geq 0}^{r_{(\ell-1)} \times r_{(\ell)}} \).

» Multi-HNTF Model (Take 2)

This model learns

\[ \mathbf{X} \approx \begin{bmatrix} \mathbf{X}_1^{(0)}, \mathbf{X}_2^{(0)}, \ldots, \mathbf{X}_k^{(0)} \end{bmatrix} \approx \begin{bmatrix} \mathbf{X}_1^{(1)}, \mathbf{X}_2^{(1)}, \ldots, \mathbf{X}_k^{(1)} \end{bmatrix} \approx \ldots \approx \begin{bmatrix} \mathbf{X}_1^{(L-1)}, \mathbf{X}_2^{(L-1)}, \ldots, \mathbf{X}_k^{(L-1)} \end{bmatrix} \]

where

\[ \mathbf{X}_j^{(\ell+1)} = \mathbf{X}_j^{(\ell)} \mathbf{W}^{(\ell)}, \]

and \( \mathbf{W}^{(\ell)} \in \mathbb{R}_{\geq 0}^{r^{(\ell-1)} \times r^{(\ell)}} \).

A single hierarchical relationship for all modes!

Training Process

1: **procedure** \textsc{Multi-HNTF}(\(\mathbf{X}\))
2: \(\{\mathbf{X}_i^{(0)}\}_{i=1}^k \leftarrow \text{NCPD}(\mathbf{X}, r_0)\)
3: for \(\ell = 0 \ldots L\) do
4: \(\mathbf{W}^{(\ell)} \leftarrow \arg\min_{\mathbf{W} \in \mathbb{R}^{r_{\ell} \times r_{\ell+1}}} \|\mathbf{X} - [\mathbf{X}_1^{(\ell)} \mathbf{W}, \ldots, \mathbf{X}_k^{(\ell)} \mathbf{W}]\|\)
5: for \(i = 0 \ldots k\) do
6: \(\mathbf{X}_i^{(\ell+1)} = \mathbf{X}_i^{(\ell)} \mathbf{W}^{(\ell)}\)
» Training Process

1: procedure \textsc{Multi-HNTF}(\mathbf{X})
2: \{\mathbf{X}_i^{(0)}\}_{i=1}^k \leftarrow \text{NCPD}(\mathbf{X}, r_0)
3: \text{for } \ell = 0 \ldots \mathcal{L} \text{ do}
4: \quad \mathbf{W}^{(\ell)} \leftarrow \text{argmin}_{\mathbf{W} \in \mathbb{R}^{r_\ell \times r_{\ell+1}}} \| \mathbf{X} - \left[ \mathbf{X}_1^{(\ell)} \mathbf{W}, \ldots, \mathbf{X}_k^{(\ell)} \mathbf{W} \right] \|
5: \quad \text{for } i = 0 \ldots k \text{ do}
6: \quad \mathbf{X}_i^{(\ell+1)} = \mathbf{X}_i^{(\ell)} \mathbf{W}^{(\ell)}

\checkmark Can be approximated via NMF method on each mode with averaging of learned \(\mathbf{W}\) matrix across modes.
» Training Process

1: procedure Multi-HNTF(\(X\))
2: \(\{X^{(0)}_i\}_{i=1}^k \leftarrow \text{NCPD}(X, r_0)\)
3: for \(\ell = 0 \ldots L\) do
4: \(W^{(\ell)} \leftarrow \text{argmin}_{W \in \mathbb{R}^{r_\ell \times (r_\ell + 1)}} \|X - [X_1^{(\ell)} W, \ldots, X_k^{(\ell)} W]\|\)
5: for \(i = 0 \ldots k\) do
6: \(X^{(\ell+1)}_i = X^{(\ell)}_i W^{(\ell)}\)

▷ Can be approximated via NMF method on each mode with averaging of learned \(W\) matrix across modes.

▷ Could/should also be trained in a neural network framework.
Experiments
The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.
The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.

<table>
<thead>
<tr>
<th>Method</th>
<th>$r_0 = 7$</th>
<th>$r_1 = 4$</th>
<th>$r_2 = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-HNTF</td>
<td>0.454</td>
<td>0.548</td>
<td>0.721</td>
</tr>
<tr>
<td>Neural HNCPD [Vendrow, et. al.]</td>
<td>0.454</td>
<td>0.508</td>
<td>0.714</td>
</tr>
<tr>
<td>Standard HNCPD [Vendrow, et. al.]</td>
<td>0.454</td>
<td>0.612</td>
<td>0.892</td>
</tr>
<tr>
<td>HNTF-1 [Cichocki, et. al.]</td>
<td>0.454</td>
<td>0.576</td>
<td>0.781</td>
</tr>
<tr>
<td>HNTF-2 [Cichocki, et. al.]</td>
<td>0.454</td>
<td>0.587</td>
<td>0.765</td>
</tr>
<tr>
<td>HNTF-3 [Cichocki, et. al.]</td>
<td>0.454</td>
<td>0.560</td>
<td>0.747</td>
</tr>
</tbody>
</table>
The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.

Projections of tensor approximation at each layer of Multi-HNTF.

Relative reconstruction error.

<table>
<thead>
<tr>
<th>Method</th>
<th>$r_0 = 7$</th>
<th>$r_1 = 4$</th>
<th>$r_2 = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-HNTF</td>
<td>0.454</td>
<td>0.548</td>
<td>0.721</td>
</tr>
<tr>
<td>Neural HNCPD [Vendrow, et. al.]</td>
<td>0.454</td>
<td>0.508</td>
<td>0.714</td>
</tr>
<tr>
<td>Standard HNCPD [Vendrow, et. al.]</td>
<td>0.454</td>
<td>0.612</td>
<td>0.892</td>
</tr>
<tr>
<td>HNTF-1 [Cichocki, et. al.]</td>
<td>0.454</td>
<td>0.576</td>
<td>0.781</td>
</tr>
<tr>
<td>HNTF-2 [Cichocki, et. al.]</td>
<td>0.454</td>
<td>0.587</td>
<td>0.765</td>
</tr>
<tr>
<td>HNTF-3 [Cichocki, et. al.]</td>
<td>0.454</td>
<td>0.560</td>
<td>0.747</td>
</tr>
</tbody>
</table>
Political Twitter Data

- A data set of tweets sent by political candidates during the 2016 election season
- We subset the tweets from eight politicians, four Republicans and four Democrats:
  
  Hillary Clinton, Tim Kaine, Martin O'Malley, Bernie Sanders, Ted Cruz, John Kasich, Marco Rubio, and Donald Trump.
### Political Twitter Data

#### Rank 8 Topics

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
<th>Topic 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>trump</td>
<td>senate</td>
<td>martinomalley</td>
<td>berniesanders</td>
</tr>
<tr>
<td>hillary</td>
<td>florida</td>
<td>hillaryclinton</td>
<td>people</td>
</tr>
<tr>
<td>donald</td>
<td>zika</td>
<td>realdonaldtrump</td>
<td>bernie</td>
</tr>
<tr>
<td>president</td>
<td>venezuela</td>
<td>campaigning</td>
<td>must</td>
</tr>
<tr>
<td>timkaine</td>
<td>nicolas maduro</td>
<td>maryland</td>
<td>change</td>
</tr>
</tbody>
</table>

#### Rank 4 Topics

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>trump</td>
<td>tedcruz</td>
</tr>
<tr>
<td>hillary</td>
<td>cruz</td>
</tr>
<tr>
<td>vote</td>
<td>ted</td>
</tr>
<tr>
<td>people</td>
<td>internet</td>
</tr>
<tr>
<td>berniesanders</td>
<td>choosercruz</td>
</tr>
</tbody>
</table>

#### Rank 2 Topics

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>johnkasich</td>
<td>crooked</td>
</tr>
<tr>
<td>kasich</td>
<td>hillary</td>
</tr>
<tr>
<td>ohio</td>
<td>thank</td>
</tr>
<tr>
<td>john</td>
<td>great</td>
</tr>
<tr>
<td>gov</td>
<td>realdonaldtrump</td>
</tr>
</tbody>
</table>

#### Method

- **r**
  - \( r_0 = 8 \)
  - \( r_1 = 4 \)
  - \( r_2 = 2 \)

- **Multi-HNTF**
  - 0.834 0.887 0.920

- **Neural HNCPD [Vendrow, et. al.]**
  - 0.834 0.883 0.918

- **Standard HNCPD [Vendrow, et. al.]**
  - 0.834 0.889 0.919

- **Standard NCPD**
  - 0.834 0.931 0.950

- **HNTF-1 [Cichocki, et. al.]**
  - 0.834 0.890 0.927

- **HNTF-2 [Cichocki, et. al.]**
  - 0.834 0.909 0.956

- **HNTF-3 [Cichocki, et. al.]**
  - 0.834 0.895 0.942

---

[19/30]
**Political Twitter Data**

<table>
<thead>
<tr>
<th>Rank 8 Topics</th>
<th>Rank 4 Topics</th>
<th>Rank 2 Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topic 1</strong></td>
<td><strong>Topic 2</strong></td>
<td><strong>Topic 3</strong></td>
</tr>
<tr>
<td>trump</td>
<td>senate</td>
<td>martinomalley</td>
</tr>
<tr>
<td>hillary</td>
<td>florida</td>
<td>hillaryclinton</td>
</tr>
<tr>
<td>donald</td>
<td>zika</td>
<td>realdonaldtrump</td>
</tr>
<tr>
<td>president</td>
<td>venezuela</td>
<td>campaigning</td>
</tr>
<tr>
<td>timkaine</td>
<td>nicolasmaduro</td>
<td>maryland</td>
</tr>
<tr>
<td><strong>Topic 5</strong></td>
<td><strong>Topic 6</strong></td>
<td><strong>Topic 7</strong></td>
</tr>
<tr>
<td>tdecruz</td>
<td>johnkasich</td>
<td>marcorubio</td>
</tr>
<tr>
<td>cruz</td>
<td>kasich</td>
<td>teammarco</td>
</tr>
<tr>
<td>ted</td>
<td>ohio</td>
<td>vote</td>
</tr>
<tr>
<td>internet</td>
<td>john</td>
<td>flsen</td>
</tr>
<tr>
<td>choosescruz</td>
<td>gov</td>
<td>click</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>$r_0 = 8$</th>
<th>$r_1 = 4$</th>
<th>$r_2 = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-HNTF</td>
<td>0.834</td>
<td>0.887</td>
<td>0.920</td>
</tr>
<tr>
<td>Neural HNCNPD [Vendrow, et. al.]</td>
<td>0.834</td>
<td>0.883</td>
<td>0.918</td>
</tr>
<tr>
<td>Standard HNCNPD [Vendrow, et. al.]</td>
<td>0.834</td>
<td>0.889</td>
<td>0.919</td>
</tr>
<tr>
<td>Standard NCPD</td>
<td>0.834</td>
<td>0.931</td>
<td>0.950</td>
</tr>
<tr>
<td>HNTF-1 [Cichocki, et. al.]</td>
<td>0.834</td>
<td>0.890</td>
<td>0.927</td>
</tr>
<tr>
<td>HNTF-2 [Cichocki, et. al.]</td>
<td>0.834</td>
<td>0.909</td>
<td>0.956</td>
</tr>
<tr>
<td>HNTF-3 [Cichocki, et. al.]</td>
<td>0.834</td>
<td>0.895</td>
<td>0.942</td>
</tr>
</tbody>
</table>
20 Newsgroups Data
» 20 Newsgroups Data

Reconstruction loss and classification accuracy at the second layer of two layer Multi-HNTF and HNMF on the 20 newsgroup data set.

<table>
<thead>
<tr>
<th>Method</th>
<th>Recon Loss</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-HNTF</td>
<td>30.81</td>
<td>30.91</td>
</tr>
<tr>
<td>HNMF</td>
<td>30.82</td>
<td>31.45</td>
</tr>
</tbody>
</table>
Backpropagation
Hierarchical Tensor Decompositions

Hierarchical NCPD

Multi-HNTF
Hierarchical Tensor Decompositions

Hierarchical NCPD

Multi-HNTF

Devastating error propagation through layers!
Reminder

**Neural Network**: Learn weights $W^{(1)}, W^{(2)}, \ldots, W^{(L)}$ to minimize model error

$$E(\{W^{(i)}\}) = \sum_{n=1}^{N} f(y(x_n, \{W^{(i)}\}), x_n, t_n).$$
Reminder

Neural Network: Learn weights $W^{(1)}, W^{(2)}, ..., W^{(L)}$ to minimize model error

$$E(\{W^{(i)}\}) = \sum_{n=1}^{N} f(y(x_n, \{W^{(i)}\}), x_n, t_n).$$
Reminder

Neural Network: Learn weights $W^{(1)}, W^{(2)}, \ldots, W^{(L)}$ to minimize model error

$$E(\{W^{(i)}\}) = \sum_{n=1}^{N} f(y(x_n, \{W^{(i)}\}), x_n, t_n).$$

Training:

- forward propagation:
  $$z^{(1)} = \sigma(W^{(1)}x),$$
  $$z^{(2)} = \sigma(W^{(2)}z_1),$$
  $$\ldots,$$
  $$y = \sigma(W^{(L)}z^{(L-1)})$$
Reminder

**Neural Network**: Learn weights $W^{(1)}, W^{(2)}, \ldots, W^{(L)}$ to minimize model error

$$E(\{W^{(i)}\}) = \sum_{n=1}^{N} f(y(x_n, \{W^{(i)}\}), x_n, t_n).$$

**Training**:

- **forward propagation**:
  - $z^{(1)} = \sigma(W^{(1)}x)$,
  - $z^{(2)} = \sigma(W^{(2)}z_1)$,
  - \ldots,
  - $y = \sigma(W^{(L)}z^{(L-1)})$

- **back propagation**:
  - update $\{W^{(i)}\}$ with $\nabla E(\{W^{(i)}\})$
Training via backpropagation

Neural NMF: Forward and back propagation algorithms for hNMF.


Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]
» **Training via backpropagation**

**Neural NMF:** Forward and back propagation algorithms for hNMF.

- Regard the $A$ matrices as independent variables, determine the $S$ matrices from the $A$ matrices.

---


Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]
Training via backpropagation

**Neural NMF:** Forward and back propagation algorithms for hNMF.

- Regard the $A$ matrices as independent variables, determine the $S$ matrices from the $A$ matrices.
- Define $q(X, A) := \arg\min_{S \geq 0} \|X - AS\|_F^2$ (least-squares).

---


Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]
Training via backpropagation

Neural NMF: Forward and back propagation algorithms for hNMF.

- Regard the $A$ matrices as independent variables, determine the $S$ matrices from the $A$ matrices.
- Define $q(X, A) := \arg\min_{S \geq 0} \|X - AS\|_F^2$ (least-squares).
- Pin the values of $S$ to those of $A$ by recursively setting $S^{(\ell)} := q(S^{(\ell-1)}, A^{(\ell)})$.


Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]
Training via backpropagation

**Neural NMF**: Forward and back propagation algorithms for hNMF.

- Regard the $A$ matrices as independent variables, determine the $S$ matrices from the $A$ matrices.
- Define $q(X, A) := \arg\min_{S \geq 0} \| X - AS \|_F^2$ (least-squares).
- Pin the values of $S$ to those of $A$ by recursively setting $S^{(\ell)} := q(S^{(\ell-1)}, A^{(\ell)})$.

---


Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]
Training via backpropagation

**Neural NMF:** Forward and back propagation algorithms for hNMF.

\[
X \xrightarrow{q(\cdot, A^{(0)})} S^{(0)} \xrightarrow{q(\cdot, A^{(1)})} S^{(1)}
\]

---


Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]
Training via backpropagation

**Neural NMF:** Forward and back propagation algorithms for hNMF.

Training:

\[ X \rightarrow S^{(0)} \rightarrow S^{(1)} \]

\[ q(\cdot, A^{(0)}) \rightarrow q(\cdot, A^{(1)}) \]

---


Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]
Training via backpropagation

**Neural NMF:** Forward and back propagation algorithms for hNMF.

\[ S^{(0)} = q(X, A^{(0)}) \]
\[ S^{(1)} = q(S^{(0)}, A^{(1)}) \]
\[ \ldots \]
\[ S^{(L)} = q(S^{(L-1)}, A^{(L)}) \]

- **forward propagation:**
  \[ q(\cdot, A^{(0)}) \]
  \[ q(\cdot, A^{(1)}) \]

- **back propagation:** update \( \{A^{(i)}\} \) with \( \nabla E(\{A^{(i)}\}) \)

---


Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]
Training via backpropagation

**Neural NMF:** Forward and back propagation algorithms for hNMF.

Training:

- **forward propagation:**
  \[ S^{(0)} = q(X, A^{(0)}), \]
  \[ S^{(1)} = q(S^{(0)}, A^{(1)}), \ldots, \]
  \[ S^{(L)} = q(S^{(L-1)}, A^{(L)}) \]

- **back propagation:** update \( \{A^{(i)}\} \) with \( \nabla E(\{A^{(i)}\}) \)

Apply this approach to each mode of HNCPD!

---


Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]
Neural NCPD

Train independent neural NMF models for each mode of tensor from fixed NCPD factor matrices.

Neural NCPD

Train independent neural NMF models for each mode of tensor from fixed NCPD factor matrices.

Gradient Calculation

**Theorem-ish** [Will, Zhang, Sadovnik, Gao, Vendrow, H., Molitor, Needell, 22+]

Given knowledge of the support of $q(A, X)$, the gradient $\nabla_A q(A, X)$ has a closed-form expression almost everywhere in the space of real-valued matrix pairs. This gradient expression is inherited from unconstrained least-squares.
Given knowledge of the support of $q(A, X)$, the gradient $\nabla_A q(A, X)$ has a closed-form expression almost everywhere in the space of real-valued matrix pairs. This gradient expression is inherited from unconstrained least-squares.
The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.
Synthetic Tensor

The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.

<table>
<thead>
<tr>
<th>Method</th>
<th>$r_0 = 7$</th>
<th>$r_1 = 4$</th>
<th>$r_2 = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-HNTF</td>
<td>0.454</td>
<td>0.548</td>
<td>0.721</td>
</tr>
<tr>
<td>Neural HNCPD [Vendrow, et. al.]</td>
<td>0.454</td>
<td><strong>0.508</strong></td>
<td><strong>0.714</strong></td>
</tr>
<tr>
<td>Standard HNCPD [Vendrow, et. al.]</td>
<td>0.454</td>
<td>0.612</td>
<td>0.892</td>
</tr>
<tr>
<td>HNTF-1 [Cichocki, et. al.]</td>
<td>0.454</td>
<td>0.576</td>
<td>0.781</td>
</tr>
<tr>
<td>HNTF-2 [Cichocki, et. al.]</td>
<td>0.454</td>
<td>0.587</td>
<td>0.765</td>
</tr>
<tr>
<td>HNTF-3 [Cichocki, et. al.]</td>
<td>0.454</td>
<td>0.560</td>
<td>0.747</td>
</tr>
</tbody>
</table>
Synthetic Tensor

The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.

Projections of tensor approximation at each layer of Multi-HNTF.

<table>
<thead>
<tr>
<th>Method</th>
<th>$r_0 = 7$</th>
<th>$r_1 = 4$</th>
<th>$r_2 = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-HNTF</td>
<td>0.454</td>
<td>0.548</td>
<td>0.721</td>
</tr>
<tr>
<td>Neural HNCPD [Vendrow, et. al.]</td>
<td>0.454</td>
<td><strong>0.508</strong></td>
<td><strong>0.714</strong></td>
</tr>
<tr>
<td>Standard HNCPD [Vendrow, et. al.]</td>
<td>0.454</td>
<td>0.612</td>
<td>0.892</td>
</tr>
<tr>
<td>HNTF-1 [Cichocki, et. al.]</td>
<td>0.454</td>
<td>0.576</td>
<td>0.781</td>
</tr>
<tr>
<td>HNTF-2 [Cichocki, et. al.]</td>
<td>0.454</td>
<td>0.587</td>
<td>0.765</td>
</tr>
<tr>
<td>HNTF-3 [Cichocki, et. al.]</td>
<td>0.454</td>
<td>0.560</td>
<td>0.747</td>
</tr>
</tbody>
</table>
Conclusions
Conclusions

- Multi-HNTF is a hierarchical tensor decomposition model that generalizes hierarchical NMF.
Conclusions

- Multi-HNTF is a hierarchical tensor decomposition model that generalizes hierarchical NMF.

- Model can be trained by your favorite NMF method with an additional projection step.
Conclusions

- Multi-HNTF is a hierarchical tensor decomposition model that generalizes hierarchical NMF.

- Model can be trained by your favorite NMF method with an additional projection step.

- Neural NMF and Neural NCPD can help mitigate devastating error propagation through multi-layer decomposition models.
Conclusions

- Multi-HNTF is a hierarchical tensor decomposition model that generalizes hierarchical NMF.

- Model can be trained by your favorite NMF method with an additional projection step.

- Neural NMF and Neural NCPD can help mitigate devastating error propagation through multi-layer decomposition models.

- Develop backpropagation framework for Multi-HNTF and first layer NCPD.
Thanks for listening!

Questions?


