Structured Non-Negative Matrix Factorization

Hans Laurberg
hla@es.aau.dk

Aalborg Universitet

Joint work with Lars Kai Hansen, Søren Holt Jensen, Mads G. Christensen and Mikkel N. Schmidt
Outline

1. Introduction to NMF
2. Structured NMF
   (a) Affine NMF
   (b) Instrumentation separation using NMF
Rank Reduction

Let $V$, $\hat{V}$, $W$ and $H$ be matrices.

Set up:

$$V \approx \hat{V} = WH$$

where $\hat{V}$ is a low rank approximation of $V$.

Purpose:

- Noise reduction
- Classification
- Data reduction
Music Example

Source: Smaragdis 2004
Principal Component Analysis (PCA)

Task:

\[
\hat{V} = \arg \min \left( \| V - \hat{V} \|_F^2 \right)
\]

\[
\text{rank}(\hat{V}) \leq d
\]

Solution:

\[
\hat{V} = \sum_{i=1}^{d} \lambda_i p_i q_i^T
\]

where \( \lambda_i, p_i \) and \( q_i \) are the singular triplets of \( V \).
Examples of positive $V$:

1. Images
2. Amplitude/Power Spectrums
3. Histograms

Interpretation of negative solution?
Positive Construction

The good news

1. Part based
2. Sparse
3. Understandable
Task:
Find element wise non-negative $W$ and $H$ that minimize the error function, $E(\hat{V})$.

Some Error Functions:

- $E_{Power}(\hat{V}) = \| V - \hat{V} \|_{F}^{2}$
- $E_{Sparse}(\hat{V}) = \| V - WH \|_{F}^{2} + \lambda \sum_{ij} H_{ij}$
- $E_{KL}(\hat{V}) = \sum_{ij} (V_{ij} \log \frac{V_{ij}}{\hat{V}_{ij}} - V_{ij} + \hat{V}_{ij})$
Generation model:

\[ V = \hat{V} + N, \]

where the elements in \( N \) are Gaussian IID.

\[
p(V|\hat{V}) \propto \prod_{ij} \exp \left( \frac{(V_{ij} - \hat{V}_{ij})^2}{-2\sigma^2} \right) = \exp \left( \frac{\|V - \hat{V}\|^2_F}{-2\sigma^2} \right)
\]

\( E_{\text{Power}} \) minimizes equivalent to ML
Sparse

Generation model $V = \hat{V} + N$, where the elements in $N$ are Gaussian IID and the prior of $H$ is exponential IID.

$$p(\hat{V}|V) \propto p(\hat{V})p(V|\hat{V})$$

$$\propto \exp(-\alpha \sum_{ij} H_{ij}) \exp\left(-\|V - \hat{V}\|_{F}^{2}/2\sigma^{2}\right)$$

$$\propto \exp\left(-\|V - \hat{V}\|_{F}^{2} - 2\alpha \sigma^{2} \sum_{ij} H_{ij}\right)^{(\sigma^{-2}/2)}$$

$E_{Sparse}$ minimizes equivalent to MAP
Kullback-Leibler Divergence

The $E_{KL}$ error function equivalent to PLSA
Gaussier and Goutte 2005
Ding, Li and Peng 2006

PLSA:

- ML of V using a mixture model with d mixtures.
- Each mixture consists of two independent variables.
Example - $W$ in the “Power” error function:

$$
\nabla_W E(V, \hat{V}) = \nabla_W \| V - WH \|^2_F
$$

$$
= 2V H^T - 2WHH^T
$$

Update rule:

$$
W = W \odot \frac{\nabla_W^+}{\nabla_W^-} = W \odot \frac{VH^T}{WHH^T}
$$
Critical Theoretical Issues

1. When does NMF exist.
2. Gaussian assumption vs. positivity.
3. Convergence to local minima.
1. \( V \approx \hat{V} = WH \)

2. Non-negative constraint leads to part based basis vectors.

3. There are some theoretical foundations for the NMF cost functions.

4. There are critical theoretical issues.
Structured NMF

How small changes can make NMF more useful.

1. **Affine NMF**
   (Laurberg and Hansen ICASSP 2007)

2. **Instrument separation using NMF**
   Laurberg and Schmidt (Ongoing work)
Affine NMF

Problem: An offset leads to non uniqueness.
Affine NMF model:

\[ \hat{V} = WH + w_0 1^T \]

Affine NMF cost function:

\[ E(\hat{V}) = \| V - \hat{V} \|_F^2 + \lambda \sum_{ij} H_{ij} \]
The “Swimmer Database” introduced by Donoho and Stodden 2004 to discuss the uniqueness issues.
The Swimmer Database

Two dimensional projection of the “Swimmer Database”.

Structured NMF – p. 19/32
Business Card Data Set

Photos plus ‘watermark’
Two dimensional projection of the business card data set.
Summary of Affine NMF

1. Offset in data occur in different kind of positive data.
2. If data has an offset, performance is improved if an affine method is used.
NMF and Instrument Separation
1. Let $V$ be the spectrogram of a music piece

2. Use training data (instruments playing solo) to find instrument models $W_1 \cdots W_N$

3. Estimate mixing coefficients $H_1 \cdots H_N$

4. $\hat{V} = \sum_i W_i H_i$

5. The instruments are separated by: $\hat{V}_i = W_i H_i$
Existing Instrument Separation

Model training:

- $V_{Bass} \approx W_{Bass} H_{temp1}$
- $V_{Drum} \approx W_{Drum} H_{temp2}$

Separation:

- $V_{Data} \approx \begin{bmatrix} W_{Bass} & W_{Drum} \end{bmatrix} \begin{bmatrix} H_{Bass} \\ H_{Drum} \end{bmatrix}$
New Method

- Is it possible to separate instruments without instrument models?
Joint Estimation and Separation

Separation:

\[
\begin{bmatrix}
V_{Data} & V_{Bass} & V_{Drum}
\end{bmatrix}
\approx
\begin{bmatrix}
W_{Bass} & W_{Drum}
\end{bmatrix}
\begin{bmatrix}
H_{Bass} & H_{temp1} & 0 \\
H_{Drum} & 0 & H_{temp2}
\end{bmatrix}
\]
Joint Estimation and Separation

Implementation?

No problem:

\[ W = W \odot \frac{\nabla^+ W}{\nabla^- W} \]
No Training Data

Structured NMF

\[
\begin{bmatrix}
V_{No\text{Piano}} & V_{No\text{Bass}} & V_{No\text{Drum}}
\end{bmatrix} \approx WH
\]

\[
= \begin{bmatrix}
W_{\text{Piano}} & W_{\text{Bass}} & W_{\text{Drum}}
\end{bmatrix}
\begin{bmatrix}
0 & * & * \\
* & 0 & * \\
* & * & 0
\end{bmatrix}
\]

Are zeros enough to ensure uniqueness? (yes)
Demo
1. Instrument separation is possible without solo songs if labels are known.
2. Easy to make update rule.
3. Ongoing work.
Questions

?