Iterative Matrix Completion and Topic Modeling Using Matrix and Tensor Factorizations

Lara Kassab

Advisor: Prof. Henry Adams. Committee: Prof. Bailey Fosdick, Prof. Michael Kirby, Prof. Chris Peterson.

Colorado State University, Department of Mathematics

PhD Defense, October 19, 2021

Overview

Overview

We propose three techniques that take into account underlying structure in large-scale data:

Overview

We propose three techniques that take into account underlying structure in large-scale data:

- An Iterative Method for Structured Matrix Completion

Overview

We propose three techniques that take into account underlying structure in large-scale data:

- An Iterative Method for Structured Matrix Completion
  - Sparsity-based structure in missing entries

Overview

We propose three techniques that take into account underlying structure in large-scale data:

- An Iterative Method for Structured Matrix Completion
  - Sparsity-based structure in missing entries
- Semi-supervised NMF Models for Topic Modeling in Learning Tasks

Overview

We propose three techniques that take into account underlying structure in large-scale data:

- An Iterative Method for Structured Matrix Completion
  - Sparsity-based structure in missing entries
- Semi-supervised NMF Models for Topic Modeling in Learning Tasks
  - Partial labels and word count data structure

Overview

We propose three techniques that take into account underlying structure in large-scale data:

- An Iterative Method for Structured Matrix Completion
  - Sparsity-based structure in missing entries
- Semi-supervised NMF Models for Topic Modeling in Learning Tasks
  - Partial labels and word count data structure
- Detecting Short-Lasting Topics Using Nonnegative Tensor Decomposition

Overview

We propose three techniques that take into account underlying structure in large-scale data:

- An Iterative Method for Structured Matrix Completion
  - Sparsity-based structure in missing entries
- Semi-supervised NMF Models for Topic Modeling in Learning Tasks
  - Partial labels and word count data structure
- Detecting Short-Lasting Topics Using Nonnegative Tensor Decomposition
  - Higher-order tensor structure

An Iterative Method for Structured Matrix Completion

An Iterative Method for Structured Matrix Completion

- Presented in Chapter 2 of my preliminary exam proposal.

An Iterative Method for Structured Matrix Completion

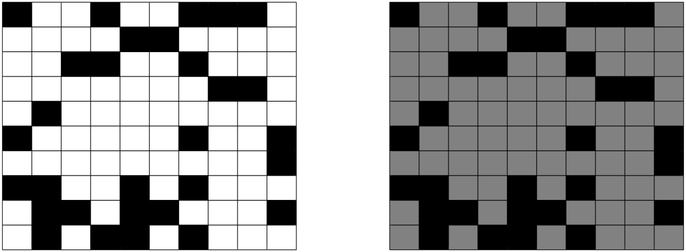
- Presented in Chapter 2 of my preliminary exam proposal.
- An Adaptation for Iterative Structured Matrix Completion. *2020 54th Asilomar Conference on Signals, Systems, and Computers*, pages 1451–1456, 2020.

An Iterative Method for Structured Matrix Completion

- Presented in Chapter 2 of my preliminary exam proposal.
- An Adaptation for Iterative Structured Matrix Completion. *2020 54th Asilomar Conference on Signals, Systems, and Computers*, pages 1451–1456, 2020.
- Journal version to appear in *Foundations of Data Science*, 2021.

Structured Matrix Completion

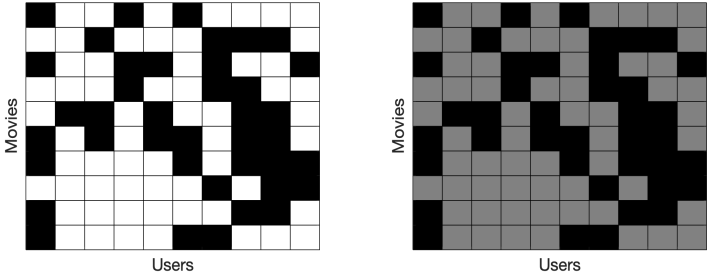
Matrix completion is the task of filling-in, or predicting, the missing entries of a partially observed matrix from a subset of known entries.



missing values   observed values   estimated values

Structured Matrix Completion

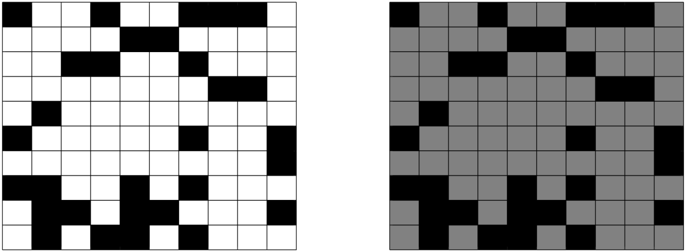
Matrix completion is the task of filling-in, or predicting, the missing entries of a partially observed matrix from a subset of known entries.



missing values observed values estimated values

Structured Matrix Completion

We consider *sparsity-based* structure in the missing entries whereby most of the missing values are close in the ℓ_0 or ℓ_1 norm sense to 0 (or more generally, to a fixed value).



missing values observed values estimated values

Structured Matrix Completion Problem

Recent work by Molitor and Needell [16] proposes solving the problem,

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \|X\|_* + \alpha \|\mathcal{P}_{\Omega^c}(X)\| \\ \text{subject to} \quad & \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M), \end{aligned} \tag{1}$$

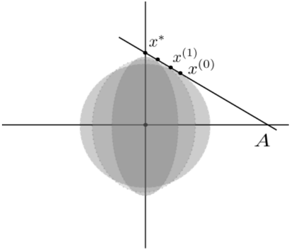
where $\alpha > 0$, $\|\cdot\|_*$ denotes the nuclear norm, $\|\cdot\|$ is an appropriate matrix norm, Ω is the set of observed entries, and $\mathcal{P}_\Omega$ denotes the sampling operator.

IRLS for Sparse Recovery

Iteratively reweighted least squares (IRLS) algorithm for sparse vector recovery [7],

$$\begin{array}{ll} \text{minimize} & \|x\|_{\ell_2(w^2)} \\ \text{subject to} & Ax = b \end{array}$$

$$x^{k+1} = \underset{x}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} w_i^k x_i^2 \; : \; Ax = b \right\} \text{ where } w^k = (|x_i^k|^2 + \epsilon_k^2)^{p/2-1}$$



IRLS for Low-rank Matrix Recovery

In [15], Mohan and Fazel propose a family of IRLS algorithms for matrix rank minimization.

The k-th iteration is given by

$$X^{k+1} = \underset{X}{\operatorname{argmin}} \left\{ \|(W_p^k)^{1/2} X\|_F^2 : \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M) \right\}, \tag{2}$$

where $W_p^k = ((X^k)^\top X^k + \gamma^k I)^{\frac{p}{2}-1}$.

Main Contribution

Low-Rank Matrix Completion	Unstructured	Structured
Non-Iterative	E.g NNM Candès and Recht	E.g. Structured NNM Molitor and Needell [17]
Iterative	E.g. IRLS-p Mohan and Fazel [16]	Our Work

Algorithm: Structured IRLS

Algorithm 1: Structured IRLS for Matrix Completion

input : $\mathcal{P}_\Omega$, M

initialize: $X^0 = \mathcal{P}_\Omega(M)$, $W_p^0 = I$, $w_q^0 = \mathbb{1}$

while *not converged* **do**

$$X^k = \underset{X}{\operatorname{argmin}} \left\{ \|(W_p^{k-1})^{1/2} X\|_F^2 + \alpha \|\mathcal{P}_{\Omega^c}(X)\|^2_{\ell_2(w_q^{k-1})} \;:\; \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M) \right\}$$

compute: $W_p^k = ((X^k)^\top X^k + \gamma^k I)^{\frac{p}{2}-1}$ and $w_q^k = ((\mathcal{P}_\Omega(X))^2 + \epsilon^k \mathbb{1})^{\frac{q}{2}-1}$

end

Implementation: Structured sIRLS

$$X^k = \underset{X}{\operatorname{argmin}} \{ \|(W_p^{k-1})^{1/2} X\|_F^2 + \alpha \|\mathcal{P}_{\Omega^c}(X)\|^2_{\ell_2(w_q^{k-1})} \ : \ \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M) \ \}$$

Algorithm 2: Structured sIRLS for Matrix Completion

input: $\mathcal{P}_\Omega$, M,r

initialize: $X^0 = \mathcal{P}_\Omega(M)$

while *not converged* **do**

- **perform :** take k_s steps promoting sparsity
- **update :** update the weights promoting low-rankness
- **perform :** take k_l steps promoting low-rankness
- **update :** update the weights promoting sparsity

end

Exact Matrix Recovery: 1000×1000 rank 10

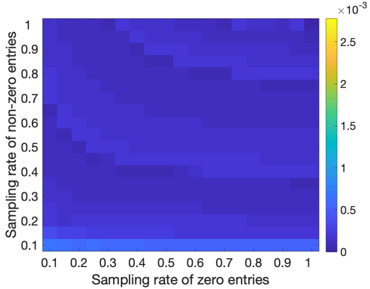


Figure: The plots display (left) the average relative error for sIRLS $\|\mathbf{M} - \tilde{\mathbf{X}}\|_F / \|\mathbf{M}\|_F$, and (right) the average relative error for Structured sIRLS $\|\mathbf{M} - \hat{\mathbf{X}}\|_F / \|\mathbf{M}\|_F$.

Exact Matrix Recovery: 1000×1000 rank 10

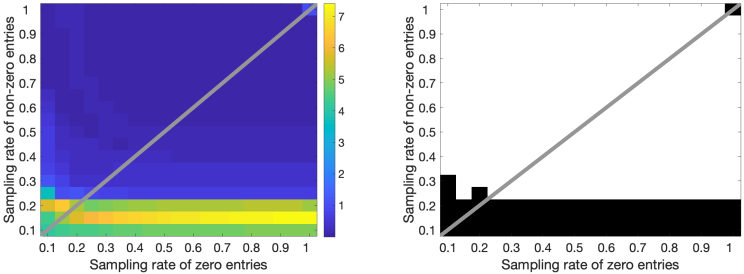


Figure: The plots display (left) the average ratio $\|\mathbf{M} - \hat{\mathbf{X}}\|_F / \|\mathbf{M} - \tilde{\mathbf{X}}\|_F$ (Structured sIRLS/sIRLS), and (right) the binned average ratio where white means the average ratio is strictly less than 1, and black otherwise.

Exact Matrix Recovery: 100×100 rank 10

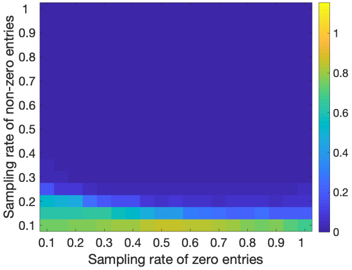


Figure: The plots display (left) the average relative error for sIRLS $\|\mathbf{M} - \tilde{\mathbf{X}}\|_F / \|\mathbf{M}\|_F$, and (right) the average relative error for Structured sIRLS $\|\mathbf{M} - \hat{\mathbf{X}}\|_F / \|\mathbf{M}\|_F$.

Exact Matrix Recovery: 100×100 rank 10

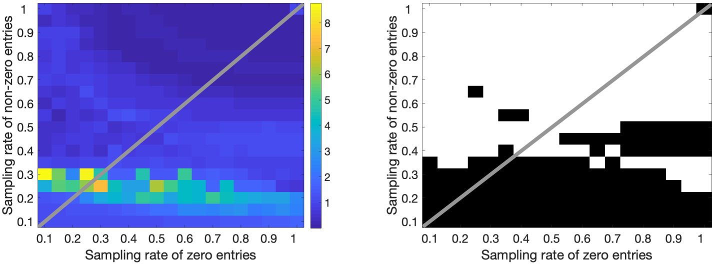


Figure: The plots display (left) the average ratio $\|\mathbf{M} - \hat{\mathbf{X}}\|_F / \|\mathbf{M} - \tilde{\mathbf{X}}\|_F$ (Structured sIRLS/sIRLS), and (right) the binned average ratio where white means the average ratio is strictly less than 1, and black otherwise.

Exact Matrix Recovery: 30×30 rank 7

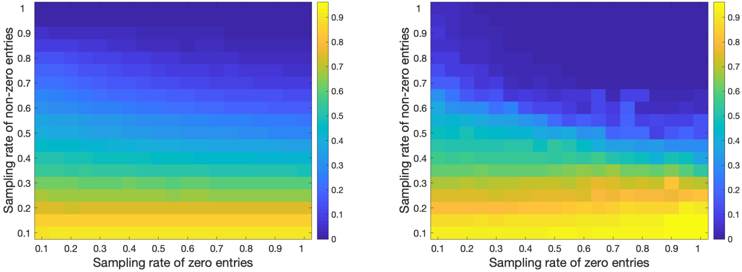


Figure: The plots display (left) the average relative error for Structured NNM $\|\mathbf{M} - \bar{\mathbf{X}}\|_F / \|\mathbf{M}\|_F$, and (right) the average relative error for Structured sIRLS $\|\mathbf{M} - \hat{\mathbf{X}}\|_F / \|\mathbf{M}\|_F$.

Exact Matrix Recovery: 30×30 rank 7

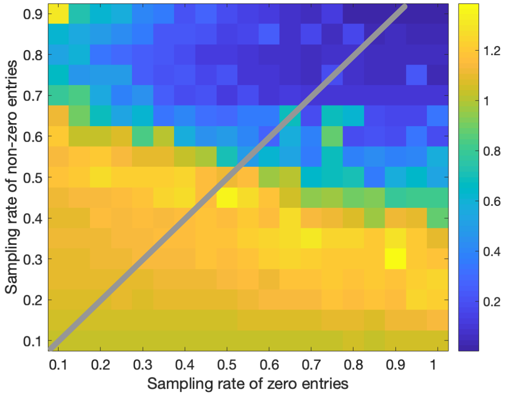


Figure: The plot displays the average ratio $\|\mathbf{M} - \hat{\mathbf{X}}\|_F / \|\mathbf{M} - \bar{\mathbf{X}}\|_F$ (Structured sIRLS/Structured NNM), when the sampling rate of non-zero entries is at most 0.90.

Matrix Completion with Noise: 100×100 rank 3

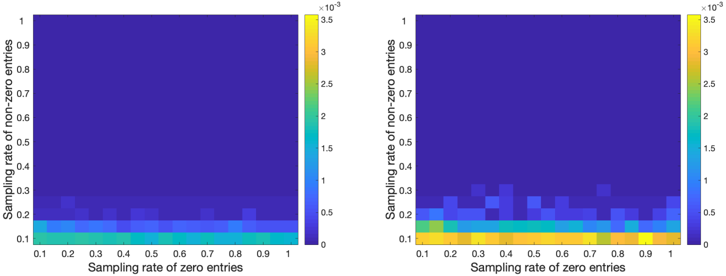


Figure: We consider twenty 100×100 random matrices of rank 3 with noise parameter $\epsilon = 10^{-4}$. The plots display (left) the average relative error for sIRLS $\|\mathbf{B} - \tilde{\mathbf{X}}\|_F / \|\mathbf{B}\|_F$, and (right) the average relative error for Structured sIRLS $\|\mathbf{B} - \hat{\mathbf{X}}\|_F / \|\mathbf{B}\|_F$.

Summary

- We adapt an iterative algorithm for low-rank matrix completion to take into account sparsity-based structure in unobserved entries by adjusting the IRLS-p algorithm studied in [15].

Summary

- We adapt an iterative algorithm for low-rank matrix completion to take into account sparsity-based structure in unobserved entries by adjusting the IRLS-p algorithm studied in [15].
- We also present an iterative gradient-projection-based implementation [11] of the algorithm that can handle large-scale matrices.

Summary

- We adapt an iterative algorithm for low-rank matrix completion to take into account sparsity-based structure in unobserved entries by adjusting the IRLS-p algorithm studied in [15].
- We also present an iterative gradient-projection-based implementation [11] of the algorithm that can handle large-scale matrices.
- We present numerical experiments showcasing improved performance of the proposed method compared to the standard IRLS algorithm [15] in structured settings on hard and noisy problems.

Summary

- We adapt an iterative algorithm for low-rank matrix completion to take into account sparsity-based structure in unobserved entries by adjusting the IRLS-p algorithm studied in [15].
- We also present an iterative gradient-projection-based implementation [11] of the algorithm that can handle large-scale matrices.
- We present numerical experiments showcasing improved performance of the proposed method compared to the standard IRLS algorithm [15] in structured settings on hard and noisy problems.
- We also show comparable performance with Structured NNM [16] for structured (small) matrices.

Future Work

- Extending the theoretical results for Structured IRLS to more general settings.

Future Work

- Extending the theoretical results for Structured IRLS to more general settings.
- Extending sparsity-based structure in the missing entries to a more general notion of structure, whereby the probability that an entry is observed or not may depend on more than just the value of that entry.

Future Work

- Extending the theoretical results for Structured IRLS to more general settings.
- Extending sparsity-based structure in the missing entries to a more general notion of structure, whereby the probability that an entry is observed or not may depend on more than just the value of that entry.
- Extending Structured IRLS methods to higher-order tensors with certain underlying low-rank and sparsity structures.

Semi-supervised NMF Models for Topic Modeling in Learning Tasks

Semi-supervised NMF Models for Topic Modeling in Learning Tasks

- This work is new and was not presented in my preliminary exam proposal.

Semi-supervised NMF Models for Topic Modeling in Learning Tasks

- This work is new and was not presented in my preliminary exam proposal.
- Semi-supervised Nonnegative Matrix Factorization for Document Classification. *arXiv preprint arXiv:2010.07956*, 2020.

Semi-supervised NMF Models for Topic Modeling in Learning Tasks

- This work is new and was not presented in my preliminary exam proposal.
- Semi-supervised Nonnegative Matrix Factorization for Document Classification. *arXiv preprint arXiv:2010.07956*, 2020.
- Selected for presentation at *2021 55th Asilomar Conference on Signals, Systems, and Computers*, 2021.

Topic Modeling

Topic modeling is an unsupervised machine learning technique used to reveal latent themes from large datasets.

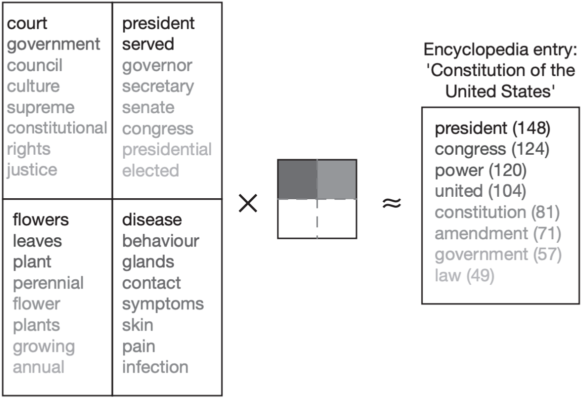
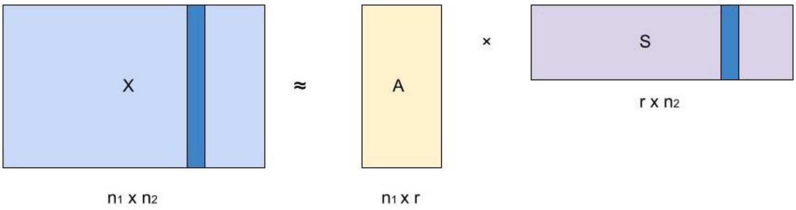


Figure: Topic modeling on 30,991 articles from the Grolier encyclopedia [13].

Nonnegative Matrix Factorization (NMF)

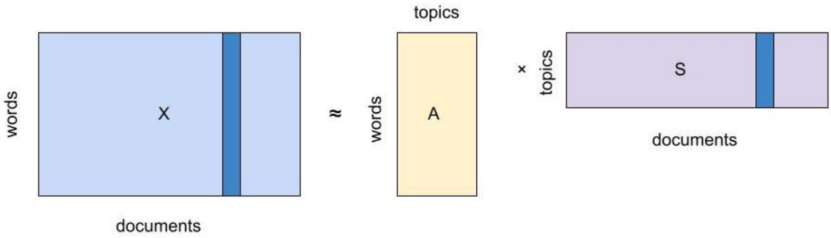
Given a matrix $X \in \mathbb{R}_{\geq 0}^{n_1 \times n_2}$ and a desired dimension r, the goal is to decompose $X \approx A \cdot S$ into two low-dimensional matrices [5, 12, 13, 27]:



$$\underset{\mathbf{A} \geq 0, \mathbf{S} \geq 0}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{AS}\|_F^2 \text{ and } \underset{\mathbf{A} \geq 0, \mathbf{S} \geq 0}{\operatorname{argmin}} D(\mathbf{X} \| \mathbf{AS})$$

Nonnegative Matrix Factorization (NMF)

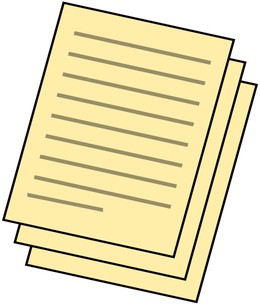
Given a matrix $X \in \mathbb{R}_{\geq 0}^{n_1 \times n_2}$ and a desired dimension r, the goal is to decompose $X \approx A \cdot S$ into two low-dimensional matrices [5, 12, 13, 27]:

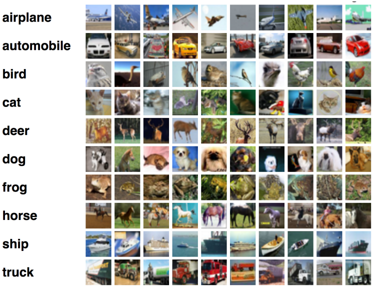


$$\underset{\mathbf{A}\geq 0, \mathbf{S}\geq 0}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{AS}\|_F^2 \quad \text{and} \quad \underset{\mathbf{A}\geq 0, \mathbf{S}\geq 0}{\operatorname{argmin}} D(\mathbf{X}\|\mathbf{AS})$$

High-Dimensional Data

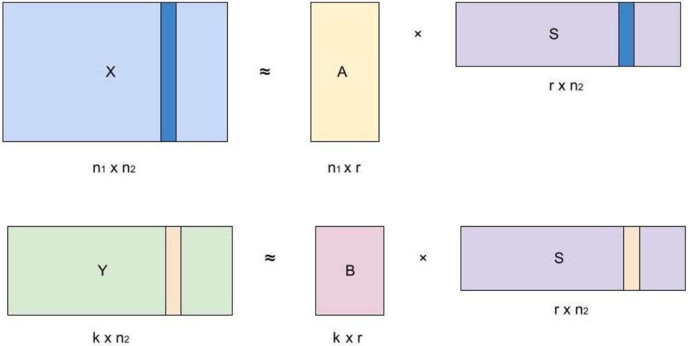
Frequently, one is faced with the problem of performing a (semi-)supervised learning task on extremely *high-dimensional data*.





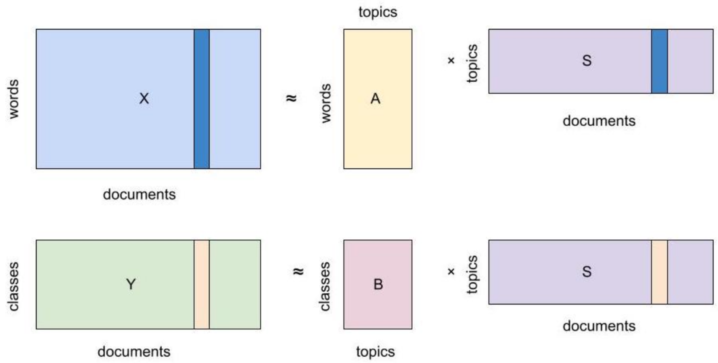
Semi-supervised NMF (SSNMF)

SSNMF, proposed in [14], is a modification of NMF to jointly incorporate a data matrix and a (partial) class label matrix:



Semi-supervised NMF (SSNMF)

SSNMF, proposed in [14], is a modification of NMF to jointly incorporate a data matrix and a (partial) class label matrix:



Semi-supervised NMF

Given a data matrix $\mathbf{X} \in \mathbb{R}_{\geq 0}^{n_1 \times n_2}$ and a class label matrix $\mathbf{Y} \in \mathbb{R}_{\geq 0}^{k \times n_2}$, SSNMF is defined by [14]:

$$\underset{\mathbf{A},\mathbf{S},\mathbf{B}\geq 0}{\operatorname{argmin}} \underbrace{\|\mathbf{X} - \mathbf{AS}\|_F^2}_{\text{Reconstruction Error}} + \lambda \underbrace{\|\mathbf{Y} - \mathbf{BS}\|_F^2}_{\text{Classification Error}}, \tag{3}$$

Multiplicative updates have been developed for SSNMF for the Frobenius norm.

Proposed SSNMF Models

In this work, we define models with different error functions applied to the reconstruction and supervision factorization terms as

$$\underset{\mathbf{A},\mathbf{S},\mathbf{B}\geq 0}{\operatorname{argmin}} \underbrace{R(\mathbf{W}\odot\mathbf{X},\mathbf{W}\odot\mathbf{AS})}_{\text{Reconstruction Error}} + \lambda \underbrace{S(\mathbf{L}\odot\mathbf{Y},\mathbf{L}\odot\mathbf{BS})}_{\text{Supervision Error}} \tag{4}$$

denoted by $(R(\cdot,\cdot), S(\cdot,\cdot))$-SSNMF where $R(\cdot,\cdot)$ and $S(\cdot,\cdot)$ are the error functions applied to the reconstruction term and supervision term, respectively.

Proposed SSNMF Models

Model	Objective
$(\|\cdot\|_F, \|\cdot\|_F)$-SSNMF [14]	$\underset{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}{\operatorname{argmin}} \|\mathbf{W} \odot (\mathbf{X} - \mathbf{AS})\|_F^2 + \lambda \|\mathbf{L} \odot (\mathbf{Y} - \mathbf{BS})\|_F^2$
$(\|\cdot\|_F, D(\cdot\|\cdot))$-SSNMF	$\underset{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}{\operatorname{argmin}} \|\mathbf{W} \odot (\mathbf{X} - \mathbf{AS})\|_F^2 + \lambda D(\mathbf{L} \odot \mathbf{Y} \| \mathbf{L} \odot \mathbf{BS})$
$(D(\cdot\|\cdot), \|\cdot\|_F)$-SSNMF	$\underset{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}{\operatorname{argmin}} D(\mathbf{W} \odot \mathbf{X} \| \mathbf{W} \odot \mathbf{AS}) + \lambda \|\mathbf{L} \odot (\mathbf{Y} - \mathbf{BS})\|_F^2$
$(D(\cdot\|\cdot), D(\cdot\|\cdot))$-SSNMF	$\underset{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}{\operatorname{argmin}} D(\mathbf{W} \odot \mathbf{X} \| \mathbf{W} \odot \mathbf{AS}) + \lambda D(\mathbf{L} \odot \mathbf{Y} \| \mathbf{L} \odot \mathbf{BS})$

Proposed SSNMF Models

- Utilizing information divergence on the data reconstruction term is a natural choice since many representations of document data correspond to counts of word patterns in the data and are naturally modelled by Poisson distributions [4, 9, 19, 20, 22, 23].

Proposed SSNMF Models

- Utilizing information divergence on the data reconstruction term is a natural choice since many representations of document data correspond to counts of word patterns in the data and are naturally modelled by Poisson distributions [4, 9, 19, 20, 22, 23].
- Further, our proposed models differ in the classification framework proposed which does not necessarily rely on SVM for linear classification.

Proposed SSNMF Models

- Utilizing information divergence on the data reconstruction term is a natural choice since many representations of document data correspond to counts of word patterns in the data and are naturally modelled by Poisson distributions [4, 9, 19, 20, 22, 23].
- Further, our proposed models differ in the classification framework proposed which does not necessarily rely on SVM for linear classification.
- We further provide analysis on the topics learned for the classification task, where the choice of rank is not necessarily the same as the number of classes.

Maximum Likelihood Estimation

Suppose that the observed data $\mathbf{X}$ and supervision information $\mathbf{Y}$ have entries given as the sum of random variables,

$$\mathbf{X}_{\gamma,\tau} = \sum_{i=1}^{r} x_{\gamma,i,\tau} \quad \text{and} \quad \mathbf{Y}_{\eta,\tau} = \sum_{i=1}^{r} y_{\eta,i,\tau},$$

and that the set of $\mathbf{X}_{\gamma,\tau}$ and $\mathbf{Y}_{\eta,\tau}$ are statistically independent conditional on $\mathbf{A}, \mathbf{B}$, and $\mathbf{S}$.

Maximum Likelihood Estimation

1. When $x_{\gamma,i,\tau}$ and $y_{\eta,i,\tau}$ have distributions

$$\mathcal{N}\left(x_{\gamma,i,\tau}|\mathbf{A}_{\gamma,i}\mathbf{S}_{i,\tau},\sigma_1\right) \text{ and } \mathcal{N}\left(y_{\eta,i,\tau}|\mathbf{B}_{\eta,i}\mathbf{S}_{i,\tau},\sigma_2\right)$$

respectively, the maximum likelihood estimator is

$$\underset{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}{\operatorname{argmin}} \|\mathbf{X}-\mathbf{AS}\|_F^2 + \frac{\sigma_1}{\sigma_2}\|\mathbf{Y}-\mathbf{BS}\|_F^2.$$

Maximum Likelihood Estimation

1. When $x_{\gamma,i,\tau}$ and $y_{\eta,i,\tau}$ have distributions

$$\mathcal{N}\left(x_{\gamma,i,\tau}|\mathbf{A}_{\gamma,i}\mathbf{S}_{i,\tau},\sigma_1\right) \text{ and } \mathcal{N}\left(y_{\eta,i,\tau}|\mathbf{B}_{\eta,i}\mathbf{S}_{i,\tau},\sigma_2\right)$$

respectively, the maximum likelihood estimator is

$$\operatorname*{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0} \|\mathbf{X}-\mathbf{AS}\|_F^2 + \frac{\sigma_1}{\sigma_2}\|\mathbf{Y}-\mathbf{BS}\|_F^2.$$

2. When $x_{\gamma,i,\tau}$ and $y_{\eta,i,\tau}$ have distributions

$$\mathcal{N}\left(x_{\gamma,i,\tau}|\mathbf{A}_{\gamma,i}\mathbf{S}_{i,\tau},\sigma_1\right) \text{ and } \mathcal{PO}\left(y_{\eta,i,\tau}|\mathbf{B}_{\eta,i}\mathbf{S}_{i,\tau}\right)$$

respectively, the maximum likelihood estimator is

$$\operatorname*{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0} \|\mathbf{X}-\mathbf{AS}\|_F^2 + 2r\sigma_1 D(\mathbf{Y}\|\mathbf{BS}).$$

Maximum Likelihood Estimation

3. When $x_{\gamma,i,\tau}$ and $y_{\eta,i,\tau}$ have distributions

$$\mathcal{PO}\left(x_{\gamma,i,\tau} | \mathbf{A}_{\gamma,i}\mathbf{S}_{i,\tau}\right) \text{ and } \mathcal{N}\left(y_{\eta,i,\tau} | \mathbf{B}_{\eta,i}\mathbf{S}_{i,\tau}, \sigma_2\right)$$

respectively, the maximum likelihood estimator is

$$\underset{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}{\operatorname{argmin}} D(\mathbf{X}\|\mathbf{AS}) + \frac{1}{2r\sigma_2}\|\mathbf{Y} - \mathbf{BS}\|_F^2.$$

Maximum Likelihood Estimation

3. When $x_{\gamma,i,\tau}$ and $y_{\eta,i,\tau}$ have distributions

$$\mathcal{PO}\left(x_{\gamma,i,\tau}|\mathbf{A}_{\gamma,i}\mathbf{S}_{i,\tau}\right) \text{ and } \mathcal{N}\left(y_{\eta,i,\tau}|\mathbf{B}_{\eta,i}\mathbf{S}_{i,\tau},\sigma_2\right)$$

respectively, the maximum likelihood estimator is

$$\underset{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}{\operatorname{argmin}} D(\mathbf{X}\|\mathbf{AS}) + \frac{1}{2r\sigma_2}\|\mathbf{Y}-\mathbf{BS}\|_F^2.$$

4. When $x_{\gamma,i,\tau}$ and $y_{\eta,i,\tau}$ have distributions

$$x_{\gamma,i,\tau} \sim \mathcal{PO}\left(x_{\gamma,i,\tau}|\mathbf{A}_{\gamma,i}\mathbf{S}_{i,\tau}\right) \text{ and } \mathcal{PO}\left(y_{\eta,i,\tau}|\mathbf{B}_{\eta,i}\mathbf{S}_{i,\tau}\right)$$

respectively, the maximum likelihood estimator is

$$\underset{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}{\operatorname{argmin}} D(\mathbf{X}\|\mathbf{AS}) + D(\mathbf{Y}\|\mathbf{BS}).$$

Proposed SSNMF Models

Model	Objective
$(\|\cdot\|_F, \|\cdot\|_F)$-SSNMF [14]	$\operatorname{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0} \|\mathbf{W} \odot (\mathbf{X} - \mathbf{AS})\|_F^2 + \lambda \|\mathbf{L} \odot (\mathbf{Y} - \mathbf{BS})\|_F^2$
$(\|\cdot\|_F, D(\cdot\|\cdot))$-SSNMF	$\operatorname{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0} \|\mathbf{W} \odot (\mathbf{X} - \mathbf{AS})\|_F^2 + \lambda D(\mathbf{L} \odot \mathbf{Y} \| \mathbf{L} \odot \mathbf{BS})$
$(D(\cdot\|\cdot), \|\cdot\|_F)$-SSNMF	$\operatorname{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0} D(\mathbf{W} \odot \mathbf{X} \| \mathbf{W} \odot \mathbf{AS}) + \lambda \|\mathbf{L} \odot (\mathbf{Y} - \mathbf{BS})\|_F^2$
$(D(\cdot\|\cdot), D(\cdot\|\cdot))$-SSNMF	$\operatorname{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0} D(\mathbf{W} \odot \mathbf{X} \| \mathbf{W} \odot \mathbf{AS}) + \lambda D(\mathbf{L} \odot \mathbf{Y} \| \mathbf{L} \odot \mathbf{BS})$

Multiplicative Updates

Suppose that the gradient of F with respect to variable Θ has a decomposition that is of the form:

$$\nabla_{\Theta} F = [\nabla_{\Theta} F]^{+} - [\nabla_{\Theta} F]^{-},$$

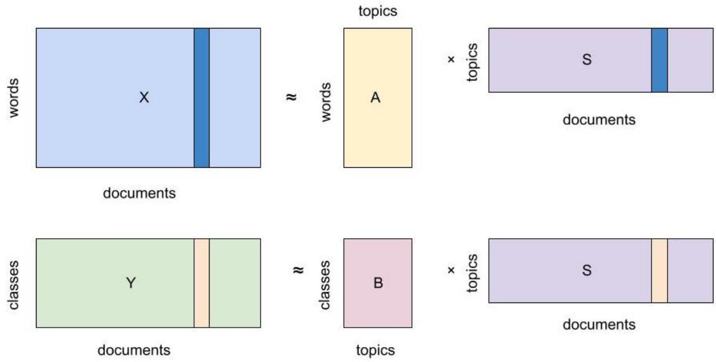
where $[\nabla_{\Theta} F]^{+} > 0$ and $[\nabla_{\Theta} F]^{-} > 0$. Then the multiplicative update for Θ has the form

$$\Theta \leftarrow \Theta \odot \frac{[\nabla_{\Theta} F]^{-}}{[\nabla_{\Theta} F]^{+}}.$$

Classification Framework

Here we describe a framework for using any of the SSNMF models for classification tasks:

$$\underset{\mathbf{A},\mathbf{S},\mathbf{B}\geq 0}{\text{argmin}} \; \underbrace{R(\mathbf{W} \odot \mathbf{X}, \mathbf{W} \odot \mathbf{AS})}_{\text{Reconstruction Error}} + \lambda \underbrace{S(\mathbf{L} \odot \mathbf{Y}, \mathbf{L} \odot \mathbf{BS})}_{\text{Supervision Error}} \tag{5}$$



Classification Framework

Here we describe a framework for using any of the SSNMF models for classification tasks:

$$\underset{\mathbf{A},\mathbf{S},\mathbf{B}\geq 0}{\operatorname{argmin}} \underbrace{R(\mathbf{W} \odot \mathbf{X}, \mathbf{W} \odot \mathbf{AS})}_{\text{Reconstruction Error}} + \lambda \underbrace{S(\mathbf{L} \odot \mathbf{Y}, \mathbf{L} \odot \mathbf{BS})}_{\text{Supervision Error}} \tag{6}$$

1. Compute $\mathbf{A}_{\text{train}}, \mathbf{B}_{\text{train}}, \mathbf{S}_{\text{train}}$ as

$$\underset{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}{\operatorname{argmin}} R(\mathbf{W}_{\text{train}} \odot \mathbf{X}_{\text{train}}, \mathbf{W}_{\text{train}} \odot \mathbf{AS}) + \lambda S(\mathbf{Y}_{\text{train}}, \mathbf{BS}).$$

Classification Framework

Here we describe a framework for using any of the SSNMF models for classification tasks:

$$\underset{\mathbf{A},\mathbf{S},\mathbf{B}\geq 0}{\operatorname{argmin}} \; \underbrace{R(\mathbf{W}\odot\mathbf{X},\mathbf{W}\odot\mathbf{AS})}_{\text{Reconstruction Error}} + \lambda \underbrace{S(\mathbf{L}\odot\mathbf{Y},\mathbf{L}\odot\mathbf{BS})}_{\text{Supervision Error}} \tag{6}$$

1. Compute $\mathbf{A}_{\text{train}}, \mathbf{B}_{\text{train}}, \mathbf{S}_{\text{train}}$ as

$$\underset{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}{\operatorname{argmin}} R(\mathbf{W}_{\text{train}}\odot\mathbf{X}_{\text{train}},\mathbf{W}_{\text{train}}\odot\mathbf{AS}) + \lambda S(\mathbf{Y}_{\text{train}},\mathbf{BS}).$$

2. Solve $\mathbf{S}_{\text{test}} = \underset{\mathbf{S}\geq 0}{\operatorname{argmin}} R(\mathbf{W}_{\text{test}}\odot\mathbf{X}_{\text{test}},\mathbf{W}_{\text{test}}\odot\mathbf{A}_{\text{train}}\mathbf{S})$.

Classification Framework

Here we describe a framework for using any of the SSNMF models for classification tasks:

$$\underset{\mathbf{A},\mathbf{S},\mathbf{B}\geq 0}{\operatorname{argmin}} \; \underbrace{R(\mathbf{W}\odot\mathbf{X},\mathbf{W}\odot\mathbf{AS})}_{\text{Reconstruction Error}} + \lambda\, \underbrace{S(\mathbf{L}\odot\mathbf{Y},\mathbf{L}\odot\mathbf{BS})}_{\text{Supervision Error}} \tag{6}$$

1. Compute $\mathbf{A}_{\text{train}}, \mathbf{B}_{\text{train}}, \mathbf{S}_{\text{train}}$ as

$$\underset{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}{\operatorname{argmin}} \; R(\mathbf{W}_{\text{train}}\odot\mathbf{X}_{\text{train}},\mathbf{W}_{\text{train}}\odot\mathbf{AS}) + \lambda S(\mathbf{Y}_{\text{train}},\mathbf{BS}).$$

2. Solve $\mathbf{S}_{\text{test}} = \underset{\mathbf{S}\geq 0}{\operatorname{argmin}} \; R(\mathbf{W}_{\text{test}}\odot\mathbf{X}_{\text{test}},\mathbf{W}_{\text{test}}\odot\mathbf{A}_{\text{train}}\mathbf{S})$.
3. Compute predicted labels as $\hat{\mathbf{Y}}_{\text{test}} = \text{label}(\mathbf{B}_{\text{train}}\mathbf{S}_{\text{test}})$, where $\text{label}(\cdot)$ assigns the largest entry of each column to 1 and all other entries to 0.

20 Newsgroups Data Experiments

Table: 20 Newsgroups and subgroups

Groups	Subgroups
Computers	graphics, mac.hardware, windows.x
Sciences	crypt(ography), electronics, space
Politics	guns, mideast
Religion	atheism, christian(ity)
Recreation	autos, baseball, hockey

20 Newsgroups Data Experiments

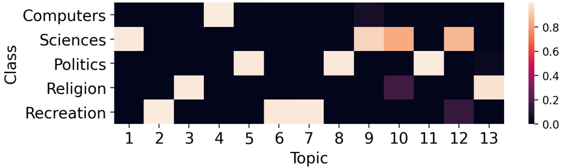


Figure: The normalized $\mathbf{B}_{\text{train}}$ matrix for the $(D(\cdot\|\cdot), \|\cdot\|_F)$ SSNMF.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12	Topic 13
would	game	god	x	would	game	players	people	would	one	israel	like	god
space	team	would	thanks	armenian	one	team	israel	chip	us	guns	anyone	people
government	car	one	anyone	one	like	car	gun	key	get	people	available	church
use	games	jesus	graphics	people	car	last	right	algorithm	could	gun	key	one
key	engine	think	know	fbi	baseball	year	government	use	like	well	probably	christians

20 Newsgroups Data Experiments

Table: Mean (and std. dev.) of test classification accuracy.

Model	Class. accuracy % (sd)
$(\|\cdot\|_F, \|\cdot\|_F)$	79.37 (0.47)
$(\|\cdot\|_F, D(\cdot\|\cdot))$	79.51 (0.38)
$(D(\cdot\|\cdot), \|\cdot\|_F)$	**81.88** (0.44)
$(D(\cdot\|\cdot), D(\cdot\|\cdot))$	81.50 (0.47)
$\|\cdot\|_F$-NMF + SVM	70.99 (2.71)
$D(\cdot\|\cdot)$-NMF + SVM	74.75 (2.50)
SVM	80.70 (0.27)
Multinomial NB	**82.28**

Summary

- We propose variants of semi-supervised nonnegative matrix factorization formulations which utilize information divergence on the data reconstruction term.

Summary

- We propose variants of semi-supervised nonnegative matrix factorization formulations which utilize information divergence on the data reconstruction term.
- We demonstrate that these models are MLE in the case of specific distributions of uncertainty assumed on the data and labels.

Summary

- We propose variants of semi-supervised nonnegative matrix factorization formulations which utilize information divergence on the data reconstruction term.
- We demonstrate that these models are MLE in the case of specific distributions of uncertainty assumed on the data and labels.
- We derive multiplicative updates for the proposed models that allow for missing data and partial supervision.

Summary

- We propose variants of semi-supervised nonnegative matrix factorization formulations which utilize information divergence on the data reconstruction term.
- We demonstrate that these models are MLE in the case of specific distributions of uncertainty assumed on the data and labels.
- We derive multiplicative updates for the proposed models that allow for missing data and partial supervision.
- We propose a classification framework for the models and demonstrate the ability of these models to perform document classification (e.g. 20 Newsgroups dataset).

Future Work

- Taking a Bayesian approach to SSNMF by assuming data-appropriate priors and performing maximum *a posteriori* estimation.

Future Work

- Taking a Bayesian approach to SSNMF by assuming data-appropriate priors and performing maximum *a posteriori* estimation.
- Forming a general framework of MLE models for exponential family distributions of uncertainty, and study the class of models where multiplicative update methods are feasible.

Future Work

- Taking a Bayesian approach to SSNMF by assuming data-appropriate priors and performing maximum *a posteriori* estimation.
- Forming a general framework of MLE models for exponential family distributions of uncertainty, and study the class of models where multiplicative update methods are feasible.
- Adapting the SSNMF algorithms for other (semi-)supervised learning tasks such as regression and generalizing such algorithms for higher-order tensors.

Detecting Short-Lasting Topics Using Nonnegative Tensor Decomposition

Detecting Short-Lasting Topics Using Nonnegative Tensor Decomposition

- Preliminary work presented in Chapter 3 of my preliminary exam proposal:

Detecting Short-Lasting Topics Using Nonnegative Tensor Decomposition

- Preliminary work presented in Chapter 3 of my preliminary exam proposal:
  - On Large-Scale Dynamic Topic Modeling with Nonnegative CP Tensor Decomposition. *Women in Data Science and Mathematics (WiSDM), Advances in Data Science, AWM-Springer series*, 2020.

Detecting Short-Lasting Topics Using Nonnegative Tensor Decomposition

- Preliminary work presented in Chapter 3 of my preliminary exam proposal:
  - On Large-Scale Dynamic Topic Modeling with Nonnegative CP Tensor Decomposition. *Women in Data Science and Mathematics (WiSDM), Advances in Data Science, AWM-Springer series*, 2020.
  - On Nonnegative CP Tensor Decomposition Robustness to Noise. *Information Theory and Applications (ITA)*, San Diego, CA, February 2020.

Detecting Short-Lasting Topics Using Nonnegative Tensor Decomposition

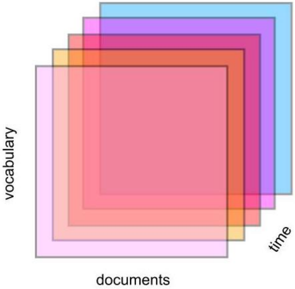
- Preliminary work presented in Chapter 3 of my preliminary exam proposal:
  - On Large-Scale Dynamic Topic Modeling with Nonnegative CP Tensor Decomposition. *Women in Data Science and Mathematics (WiSDM), Advances in Data Science, AWM-Springer series*, 2020.
  - On Nonnegative CP Tensor Decomposition Robustness to Noise. *Information Theory and Applications (ITA)*, San Diego, CA, February 2020.
- New work not presented in my preliminary proposal:

Detecting Short-Lasting Topics Using Nonnegative Tensor Decomposition

- Preliminary work presented in Chapter 3 of my preliminary exam proposal:
  - On Large-Scale Dynamic Topic Modeling with Nonnegative CP Tensor Decomposition. *Women in Data Science and Mathematics (WiSDM), Advances in Data Science, AWM-Springer series*, 2020.
  - On Nonnegative CP Tensor Decomposition Robustness to Noise. *Information Theory and Applications (ITA)*, San Diego, CA, February 2020.
- New work not presented in my preliminary proposal:
  - Detecting Short-lasting Topics Using Nonnegative Tensor Decomposition. *arXiv preprint arXiv:2010.01600*, 2021. Submitted for publication.

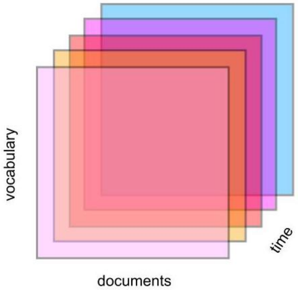
Dynamic Topic Modeling

Dynamic topic modeling investigates how topics evolve in a sequentially organized corpus of documents.



Dynamic Topic Modeling

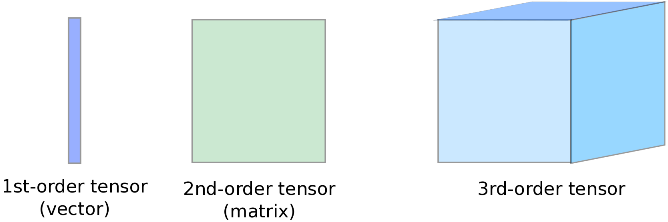
Dynamic topic modeling investigates how topics evolve in a sequentially organized corpus of documents.



A truly successful topic modeling strategy should be able to detect both long-lasting and shortly-lasting topics and clearly locate them in time.

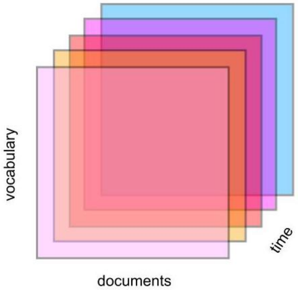
Overview on Tensors

A tensor is a multidimensional or N-way array.

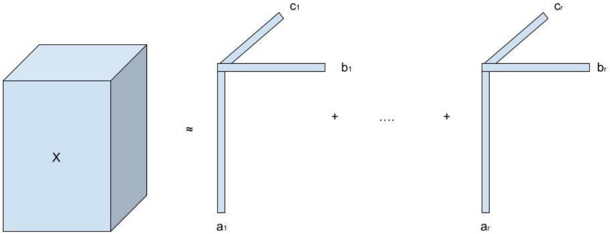


NCPD for Topic Modeling

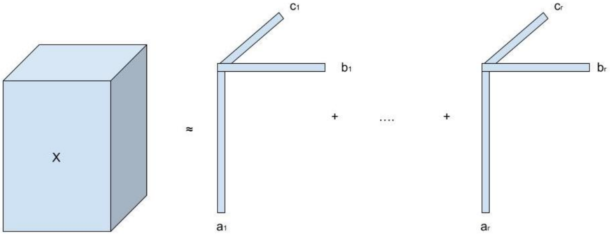
The proposed method, based on nonnegative CANDECOMP/PARAFAC tensor decomposition (NCPD) [2, 8], processes the entire 3-dimensional tensor at once.



Nonnegative CP Decomposition (NCPD)



Nonnegative CP Decomposition (NCPD)



Given a third-order tensor $\mathcal{X} \in \mathbb{R}_+^{n_1 \times n_2 \times n_3}$ and a desired dimension r, the approximate NCPD of $\mathcal{X}$ seeks $\mathbf{A} \in \mathbb{R}_+^{n_1 \times r}, \mathbf{B} \in \mathbb{R}_+^{n_2 \times r}, \mathbf{C} \in \mathbb{R}_+^{n_3 \times r}$ so that

$$\mathcal{X} \approx \sum_{\ell=1}^{r} a_\ell \otimes b_\ell \otimes c_\ell,$$

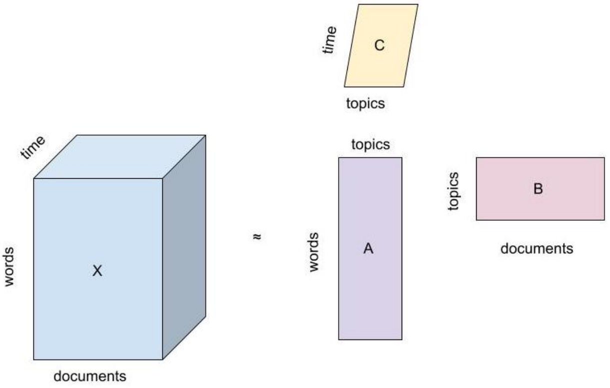
where $\otimes$ denotes the outer product and a_ℓ, b_ℓ, and c_ℓ are the columns of $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$, respectively.

Nonnegative CP Decomposition (NCPD)

We consider minimizing the *reconstruction error*,

$$\left\| \mathcal{X} - \sum_{k=1}^{r} a_k \otimes b_k \otimes c_k \right\|_F$$

among all the nonnegative vectors a_k, b_k, and c_k.



Latent Dirichlet Allocation (LDA)

LDA is a generative, probabilistic model:

- documents are probability distributions over latent topics;
- topics are probability distributions over words.

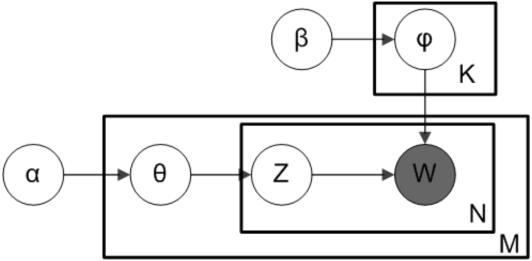


Figure: Plate notation for LDA with Dirichlet-distributed topic-word distributions. Source: Wikipedia.

Latent Dirichlet Allocation (LDA)

Algorithm 6: Generative model for latent Dirichlet allocation

for all topics $k \in [1, K]$ **do**
    sample mixture components $\phi_k \sim Dir(\beta)$
end for
for all documents $m \in [1, M]$ **do**
    sample mixture proportion $\theta_m \sim Dir(\alpha)$
    sample document length $N_m \sim Poiss(\zeta)$
    for all words $n \in [1, N_m]$ in document m **do**
        sample topic index $z_{m,n} \sim Mult(\theta_m)$
        sample term for word $w_{m,n} \sim Mult(\phi_{z_{m,n}})$
    end for
end for

Semi-synthetic 20 Newsgroups Dataset Preprocessing

Semi-synthetic 20 Newsgroups Dataset:

- The 20 Newsgroups dataset is a collection of documents divided into six groups partitioned into subjects, with a total of 20 subgroups.
- We consider only five categories: "Atheism", "Space", "Baseball", "For Sale", and "Windows X" with a total of 1040 documents.

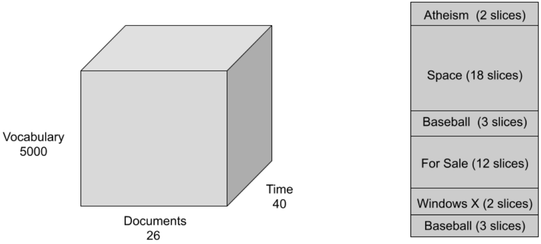


Figure: Semi-synthetic 20 Newsgroups tensor construction.

Results on 20 Newsgroups Dataset

	NMF	NCPD	LDA
1	would, like, think	space, would, like	would, like, one
2	drive, sale, offer	00, sale, drive	edu, use, window
3	space, shuttle, nasa	games, game, year	space, launch, nasa
4	00, 20, 50	god, believe, religion	new, sale, please
5	mac, hm, msu	window, widget, application	00, 50, 20

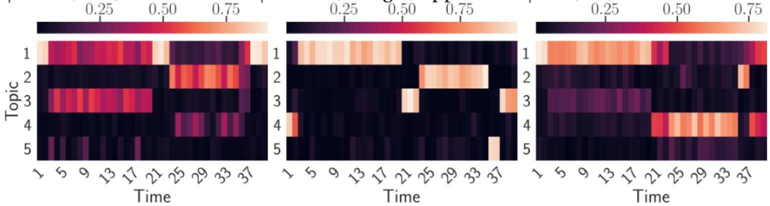


Figure: The learned topics and prevalence of each extracted topic from the semi-synthetic 20 Newsgroups dataset for (NMF, NCPD, LDA).

Twitter COVID-19 dataset Preprocessing

Twitter COVID-19 data:

- We consider Twitter text data related to the COVID-19 pandemic from Feb. 1 to May 1 of 2020 [3].
- Specifically, we consider the top 1000 retweeted English tweets from each day.

NCPD Results on COVID-19 Twitter Dataset

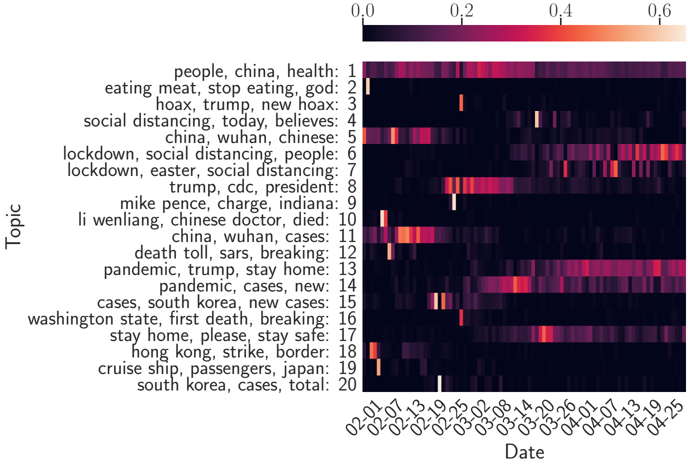


Figure: The normalized factor matrix of NCPD with rank 20. Each column of the heatmap indicates the distribution over the extracted topics for each day.

News Sources

Date	News
2 Feb	‘No Meat, No Coronavirus’ [26]
3 Feb	COVID-19 cruise ship outbreak [10]
7 Feb	Death of Dr. Li Wenliang [1]
8 Feb	COVID-19 death toll overtakes SARS [6]
18 Feb	Spike of cases in South Korea [24]
26 Feb	Mike Pence appointed to lead Coronavirus task force [21]
28 Feb	‘Trump calls Coronavirus Democrats’ “new hoax” [18]
29 Feb	First COVID-19 death in the U.S. [17]
11 Mar	WHO declares a pandemic [25]

Table: Event dates, headline summaries and references for news events relevant to identified topics.

NMF Results on COVID-19 Twitter Dataset

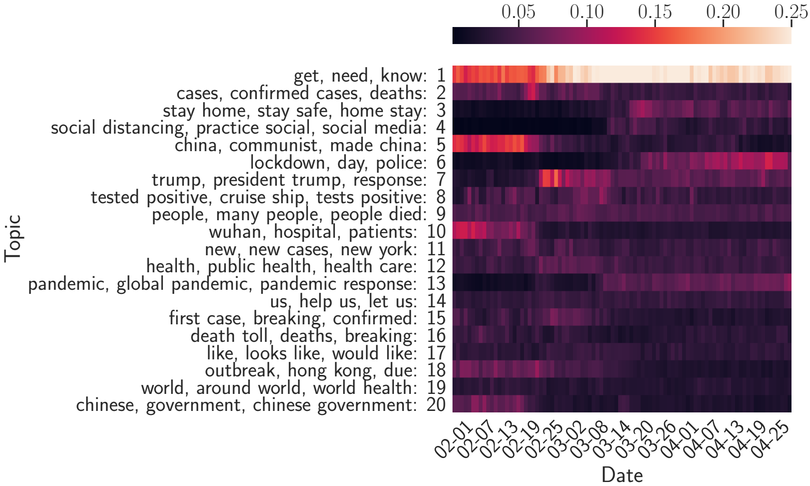


Figure: The normalized mean topic representation of tweets per day learned via NMF with rank 20.

LDA Results on COVID-19 Twitter Dataset

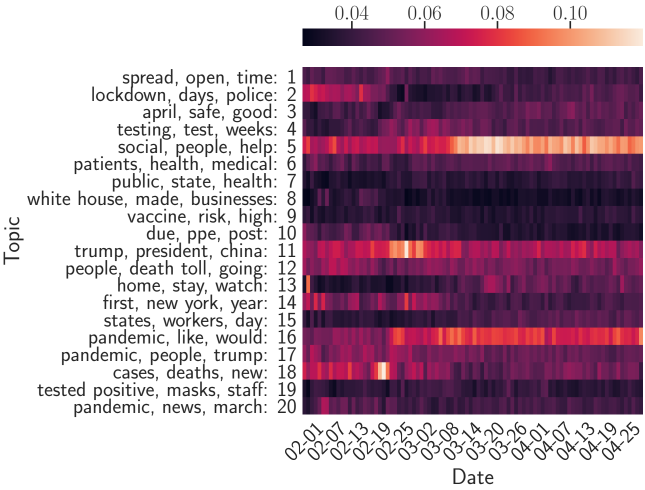


Figure: The mean topic representation of tweets per day learned via LDA with rank 20.

Million Headlines Dataset Preprocessing

Million Headlines Dataset:

- Dataset containing news headlines published from the years 2003 to 2019 sourced from the Australian news source ABC.
- We consider 700 headlines randomly selected per month with a total of 142,100 headlines in the entire dataset.

NCPD Results on News Headlines Dataset

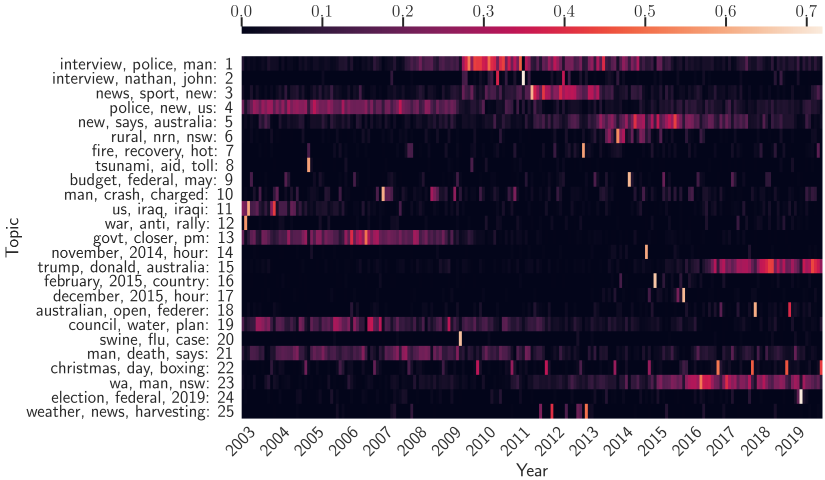


Figure: The normalized factor matrix of NCPD on the News Headlines dataset with rank 25.

NMF Results on News Headlines Dataset

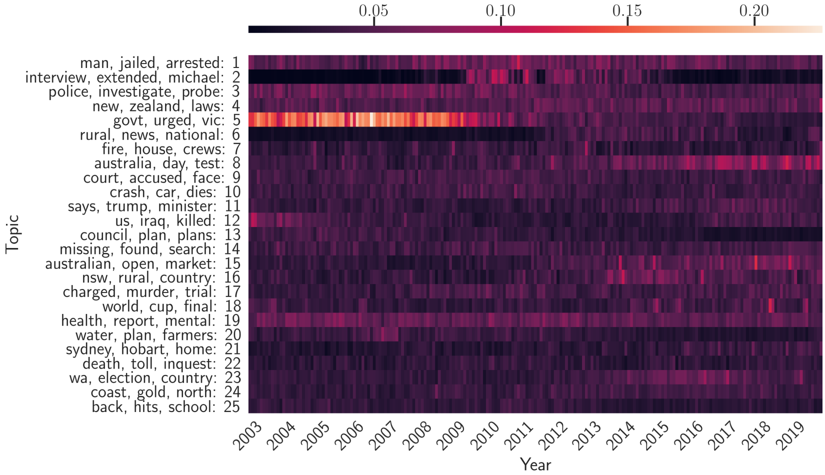


Figure: The normalized mean topic representation of headlines per month learned via NMF with rank 25.

LDA Results on News Headlines Dataset

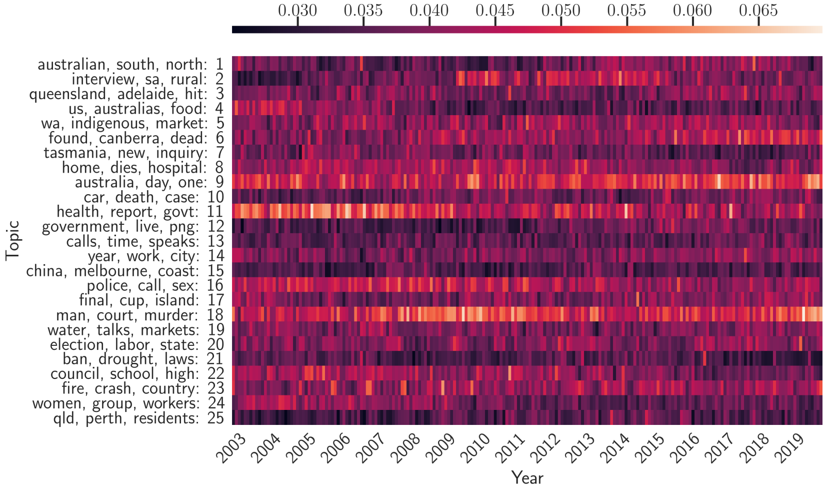


Figure: The mean topic representation of headlines per month learned via LDA with rank 25.

Summary

- We demonstrate NCPD as a dynamic topic modeling technique able to detect and accurately represent long-lasting, short-lasting, and periodic topics from temporal text data.

Summary

- We demonstrate NCPD as a dynamic topic modeling technique able to detect and accurately represent long-lasting, short-lasting, and periodic topics from temporal text data.
- NCPD performs significantly better than standard matrix-based topic modeling methods such as LDA and NMF in detecting topics with short durations.

Summary

- We demonstrate NCPD as a dynamic topic modeling technique able to detect and accurately represent long-lasting, short-lasting, and periodic topics from temporal text data.
- NCPD performs significantly better than standard matrix-based topic modeling methods such as LDA and NMF in detecting topics with short durations.
- An interesting auxiliary result of this work is the analysis of Twitter text data related to the COVID-19 pandemic [3].

Future Work

- Improving the efficiency of nonnegative tensor decompositions.

Future Work

- Improving the efficiency of nonnegative tensor decompositions.
- Controlling the average desired length of the discovered topics, thus forcing the algorithm to focus more or less on the short-term events.

Future Work

- Improving the efficiency of nonnegative tensor decompositions.
- Controlling the average desired length of the discovered topics, thus forcing the algorithm to focus more or less on the short-term events.
- Performing comparisons between matrix and tensor-based methods on different types of data (e.g. hyperspectral image data) for topic modeling and other applications.

Conclusion

Conclusion

- The first technique is an iterative method for matrix completion designed to handle large-scale datasets and take into account sparsity-based structure in the missing values to improve recovery.

Conclusion

- The first technique is an iterative method for matrix completion designed to handle large-scale datasets and take into account sparsity-based structure in the missing values to improve recovery.
- The second is a semi-supervised nonnegative matrix factorization formulation with information divergence to better model word count data and learn a low-dimensional representation that serves a (semi-)supervised machine learning task.

Conclusion

- The first technique is an iterative method for matrix completion designed to handle large-scale datasets and take into account sparsity-based structure in the missing values to improve recovery.
- The second is a semi-supervised nonnegative matrix factorization formulation with information divergence to better model word count data and learn a low-dimensional representation that serves a (semi-)supervised machine learning task.
- The third is a dynamic topic modeling technique that uses nonnegative tensor decomposition to simultaneously process all the modes of the data tensor, resulting in a more time-localized lower-dimensional representation than traditional matrix methods.

Conclusion

- The first technique is an iterative method for matrix completion designed to handle large-scale datasets and take into account sparsity-based structure in the missing values to improve recovery.
- The second is a semi-supervised nonnegative matrix factorization formulation with information divergence to better model word count data and learn a low-dimensional representation that serves a (semi-)supervised machine learning task.
- The third is a dynamic topic modeling technique that uses nonnegative tensor decomposition to simultaneously process all the modes of the data tensor, resulting in a more time-localized lower-dimensional representation than traditional matrix methods.
- Each method exploits a different structure in the data: sparsity-based structure, partial labels and word count data, and higher-order tensor structure.

Thank you for coming to my defense and for being part of this journey!

A special thank you to Henry, my committee members, friends, and family.

References I

[1] BBC. Li Wenliang: Coronavirus kills Chinese whistleblower doctor, 2020. URL https://www.bbc.com/news/world-asia-china-51403795.

[2] J Douglas Carroll and Jih-Jie Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35(3): 283–319, 1970.

[3] Emily Chen, Kristina Lerman, and Emilio Ferrara. Tracking social media discourse about the COVID-19 pandemic: Development of a public Coronavirus Twitter data set. *JMIR Public Health and Surveillance*, 6(2):e19273, 2020.

[4] Eric C Chi and Tamara G Kolda. On tensors, sparsity, and nonnegative factorizations. *SIAM Journal on Matrix Analysis and Applications*, 33(4):1272–1299, 2012.

[5] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[6] CNBC. Global death toll for new Coronavirus soars past 800 to overtake sars, 2020. URL https://www.cnbc.com/2020/02/09/coronavirus-china-death-toll-for-new-virus-exceeds-that-of-sars.html.

[7] Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, and C Sinan Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 63(1):1–38, 2010.

[8] Richard A Harshman et al. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis. 1970.

[9] Le Thi Khanh Hien and Nicolas Gillis. Algorithms for nonnegative matrix factorization with the Kullback–Leibler divergence. *Journal of Scientific Computing*, 87(3):1–32, 2021.

References II

[10] K Kakimoto, H Kamiya, T Yamagishi, T Matsui, M Suzuki, and T Wakita. Initial investigation of transmission of COVID-19 among crew members during quarantine of a cruise ship — Yokohama, Japan, February 2020. *MMWR Morb Mortal Wkly Rep*, 69: 312–313, 2020. URL `http://dx.doi.org/10.15585/mmwr.mm6911e2`.

[11] Lara Kassab, Henry Adams, and Deanna Needell. Structured IRLS code. `https://github.com/lara-kassab/structured-matrix-completion-IRLS`.

[12] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Proc. Adv. Neur. In.*, pages 556–562, 2001.

[13] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788, 1999.

[14] H. Lee, J. Yoo, and S. Choi. Semi-supervised nonnegative matrix factorization. *IEEE Signal Proc. Let.*, 17(1):4–7, 2009.

[15] Karthik Mohan and Maryam Fazel. Iterative reweighted algorithms for matrix rank minimization. *Journal of Machine Learning Research*, 13(Nov):3441–3473, 2012.

[16] Denali Molitor and Deanna Needell. Matrix completion for structured observations. In *2018 Information Theory and Applications Workshop (ITA)*, pages 1–5. IEEE, 2018.

[17] NBCNews. Washington state man becomes first U.S. death from Coronavirus, 2020. URL `https://www.nbcnews.com/news/us-news/1st-coronavirus-death-u-s-officials-say-n1145931`.

[18] NBCNews. Trump calls Coronavirus democrats' 'new hoax', 2020. URL `https://www.nbcnews.com/politics/donald-trump/trump-calls-coronavirus-democrats-new-hoax-n1145721`.

[19] John Ashworth Nelder and Robert WM Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3):370–384, 1972.

References III

[20] Duy Khuong Nguyen and Tu Bao Ho. Fast parallel randomized algorithm for nonnegative matrix factorization with KL divergence for large sparse datasets. *arXiv preprint arXiv:1604.04026*, 2016.

[21] Politico. Trump puts Pence in charge of Coronavirus response, 2020. URL `https://www.politico.com/news/2020/02/26/trump-puts-pence-in-charge-of-coronavirus-response-117790`.

[22] Anand Rajaraman and Jeffrey David Ullman. *Mining of massive datasets*. Cambridge University Press, 2011.

[23] German Rodriguez. Generalized linear models. *Lecture notes [acessado em 1 Mai 2010]. Disponível em: http://data. princeton. edu/wws509/notes*, 2001.

[24] Statista. Number of new Coronavirus (COVID-19) cases in South Korea from January 20 to August 21, 2020, 2020. URL `https://www.statista.com/statistics/1102777/south-korea-covid-19-daily-new-cases/`.

[25] WHO. Timeline of who's response to COVID-19, 2020. URL `https://www.who.int/news-room/detail/29-06-2020-covidtimeline`.

[26] Wire. 'No meat, no Coronavirus' makes no sense, 2020. URL `https://thewire.in/health/meat-eating-2019-novel-coronavirus-wuhan-bats-proteins-cattle-climate-cow-vigilantism`

[27] Z. Yang, H. Zhang, Z. Yuan, and E. Oja. Kullback-Leibler divergence for nonnegative matrix factorization. In *Proc. Int. Conf. on Artif. Neural Networks*, pages 250–257. Springer, 2011.