Adversarially-Trained Nonnegative Matrix Factorization

Ting Cai[1], Vincent Y. F. Tan[2], Cédric Févotte[3]

[1] University of Wisconsin-Madison, [2] National University of Singapore, [3] IRIT Université de Toulouse

Poster ID: 9296

Nonnegative Matrix Factorization

- Given a nonnegative data matrix $\mathbf{V} \in \mathbb{R}_+^{F\times N}$, approximate $\mathbf{V} \approx \mathbf{WH}$ where $\mathbf{W} \in \mathbb{R}_+^{F\times K}$ (basis) and $\mathbf{H} \in \mathbb{R}_+^{F\times N}$ (coefficients). To find such a decomposition, one solves

$$\min_{\mathbf{W},\mathbf{H}\geq \mathbf{0}} \|\mathbf{V} - \mathbf{WH}\|_F^2$$

where $\mathbf{A} \geq \mathbf{0}$ means $\mathbf{A}$'s entries are nonnegative.

- **Motivation**: Improve the predictive performance of NMF using adversarial training for matrix completion tasks.
- **Main Contribution**: Derive efficient algorithms for updating the adversary and $(\mathbf{W},\mathbf{H})$. Demonstrate the superior performance of adversarially-trained NMF or AT-NMF over other methods on matrix completion tasks for three benchmark datasets.

Formulation of AT-NMF

- Consider an adversary adds an *arbitrary* matrix $\mathbf{R} \in \mathbb{R}^{F\times N}$ to $\mathbf{V}$ to maximize the dirvergence between $\mathbf{V}$ and $\mathbf{WH}$, AT-NMF is formulated as

$$\min_{\mathbf{W},\mathbf{H}\geq \mathbf{0}} \max_{\mathbf{R}\in\mathcal{R}} \|\mathbf{V}+\mathbf{R} - \mathbf{WH}\|_F^2$$

where the constraint set $\mathcal{R} := \left\{\mathbf{R} : \|\mathbf{R}\|_F^2 \leq \epsilon, \mathbf{V}+\mathbf{R}\geq \mathbf{0}\right\}$.

- $\epsilon > 0$ is a constant indicating the adversary's power.
- To relax the probelm, dualize $\|\mathbf{R}\|_F^2 \leq \epsilon$ with a *Lagrange multiplier* $\lambda > 0$, AT-NMF becomes

$$\min_{\mathbf{W},\mathbf{H}\geq \mathbf{0}} \max_{\mathbf{R}:\mathbf{V}+\mathbf{R}\geq\mathbf{0}} \|\mathbf{V}+\mathbf{R} - \mathbf{WH}\|_F^2 - \lambda\|\mathbf{R}\|_F^2.$$

AT-NMF Algorithm

- **Update of R**
  - Let $\hat{\mathbf{V}} = \mathbf{WH}$, the inner maximization problem can be rewritten as a minimization problem as

$$\mathbf{R}^* = \arg\min_{\mathbf{R}:\mathbf{V}+\mathbf{R}\geq\mathbf{0}} -\|\mathbf{V}+\mathbf{R}-\hat{\mathbf{V}}\|_F^2 + \lambda\|\mathbf{R}\|_F^2$$

  - The objective decomposes into the sum of FN independent terms

$$g(\mathbf{R}) = \sum_{f,n}^{F,N}\left[-(v_{fn}+r_{fn}-\hat{v}_{fn})^2 + \lambda r_{fn}^2\right].$$

  - It suffices to minimize each term inside over r_{fn}. By re-arranging:

$$\min_{r_{fn}:v_{fn}+r_{fn}\geq 0} (\lambda-1)r_{fn}^2 - 2r_{fn}(v_{fn}-\hat{v}_{fn})$$

AT-NMF Algorithm (cont)

  - $\lambda \in [0,1]$ leads r_{fn}^* to ∞. Choose $\lambda > 1$, the update of $\mathbf{R}$ is

$$\mathbf{R} = \max\left\{\frac{\mathbf{V}-\hat{\mathbf{V}}}{\lambda-1}, -\mathbf{V}\right\}$$

- **Update of** $(\mathbf{W},\mathbf{H})$
  - Regard $\mathbf{U} := \mathbf{V}+\mathbf{R}^*$, by Majorization-Minimization (MM) algorithm,

$$\mathbf{H} \leftarrow \mathbf{H}\cdot\frac{\mathbf{W}^\top\mathbf{U}}{\mathbf{W}^\top\mathbf{WH}} \quad\text{and}\quad \mathbf{W} \leftarrow \mathbf{W}\cdot\frac{\mathbf{UH}^\top}{\mathbf{WHH}^\top}$$

- **Initialization of** $(\mathbf{W},\mathbf{H})$
  - Sample each entries independently from Half-Normal distribution ($\gamma = 1$),
  - Run 5 standard MM steps on $\mathbf{V}$ to obtain $\mathbf{W}_{\text{init}}$ and $\mathbf{H}_{\text{init}}$.
- **Stopping Criteria of AT-NMF**
  - $(\mathbf{W}^{(o,i)},\mathbf{H}^{(o,i)})$ denotes the iterate of $(\mathbf{W},\mathbf{H})$ at the o^{th} outer iteration and i^{th} inner iteration and $\hat{\mathbf{V}}^{(o,i)} := \mathbf{W}^{(o,i)}\mathbf{H}^{(o,i)}$. ε_{in} and ε_{out} are positive constants.
  - Terminate the inner optimization once the inner iteration i satisfies

$$\left\|\frac{\hat{\mathbf{V}}^{(o,i+1)}-\hat{\mathbf{V}}^{(o,i)}}{\hat{\mathbf{V}}^{(o,i)}}\right\|_F < \varepsilon_{\text{in}}$$

  - Terminate the entire optimization once the outer iteration o satisfies

$$\left\|\frac{\hat{\mathbf{V}}^{(o+1,i)}-\hat{\mathbf{V}}^{(o,i)}}{\hat{\mathbf{V}}^{(o,i)}}\right\|_F < \varepsilon_{\text{out}}$$

Synthetic Dataset

- $\alpha \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ denotes the fraction of held-out entries. $\Gamma \subset \{1,\cdots,F\}\times\{1,\cdots,N\}$ is the set of held-out entries of $\mathbf{V}$. $\hat{v}_{fn}$ is the prediction of v_{fn}. Our performance metric is the root mean-squared error (RMSE)

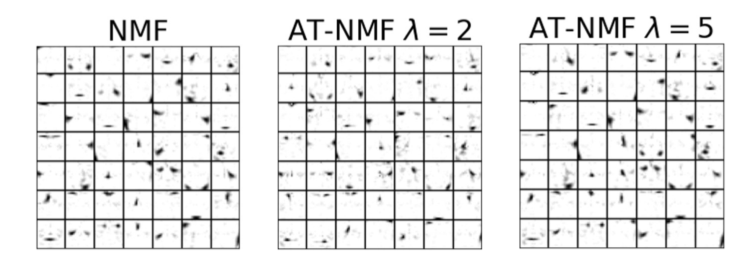
$$\text{RMSE} := \sqrt{\frac{1}{|\Gamma|}\sum_{(f,n)\in\Gamma}\left(v_{fn}-\hat{v}_{fn}\right)^2}$$

- Systhetic dataset of $F = 100$, $N = 50$, and $K = 5$.
- RMSE of Synthetic dataset

α	NMF	ANMF	AT-NMF (2)	AT-NMF (3)	AT-NMF (5)
0.3	5.37 ± 0.02	6.78 ± 0.17	5.41 ± 0.12	$\mathbf{5.11\pm0.03}$	5.20 ± 0.02
0.4	5.62 ± 0.03	6.92 ± 0.17	5.54 ± 0.08	$\mathbf{5.32\pm0.09}$	5.42 ± 0.04
0.5	6.41 ± 0.01	7.44 ± 0.09	6.27 ± 0.11	$\mathbf{6.05\pm0.03}$	6.18 ± 0.02
0.6	6.74 ± 0.02	7.61 ± 0.09	6.47 ± 0.07	$\mathbf{6.39\pm0.03}$	6.53 ± 0.02
0.7	7.30 ± 0.01	7.99 ± 0.06	7.02 ± 0.04	$\mathbf{6.94\pm0.01}$	7.10 ± 0.02
0.8	7.87 ± 0.01	8.30 ± 0.06	7.69 ± 0.04	$\mathbf{7.61\pm0.03}$	7.71 ± 0.00
0.9	8.45 ± 0.01	8.58 ± 0.06	8.44 ± 0.02	$\mathbf{8.34\pm0.02}$	8.35 ± 0.02

CBCL Dataset

- $N = 2429$ facial images with $F = 361$ pixels.
- Parts learnt when $\alpha = 0.1$



- Image Restoration by AT-NMF

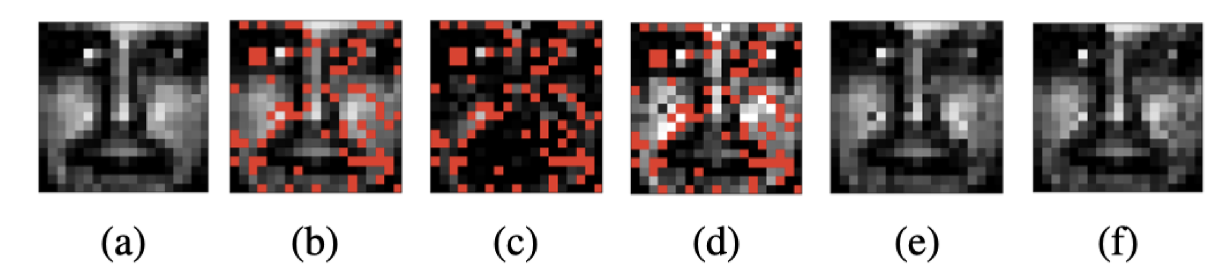
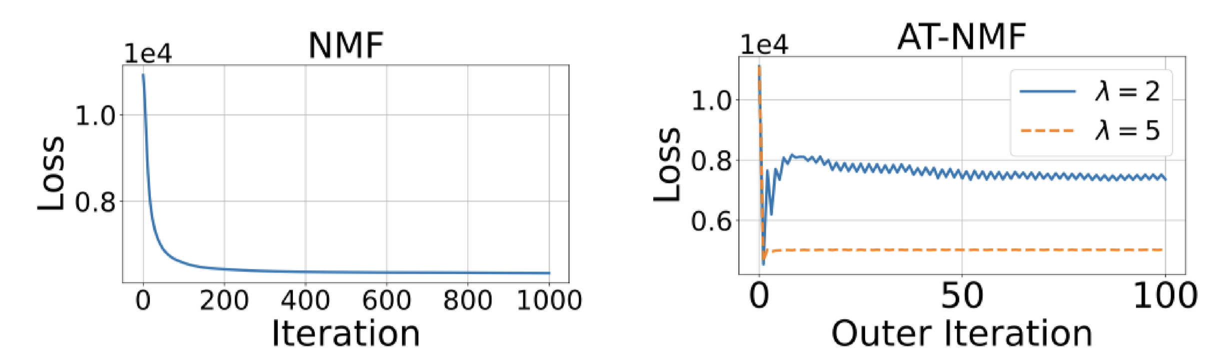


Fig. 3: (a) Original Imgage $\mathbf{V}$; (b) Masked training image (missing pixels in red) $\mathbf{V}$; (c) Adversary's added-on masked image R^*; (d) Adversarially trained masked image $\mathbf{V}+R^*$; (e) Restored image using AT-NMF; (f) Restored image using NMF when $\alpha = 0.2$ and $\lambda = 2$.

- Training losses when $\alpha = 0.5$



Hyperspectral Dataset

- It includes the Moffet and Madonna datasets with $F = 165$ and $F = 160$ respectively and $N = 2500$.
- Effect of λ on the RMSE