Fast Robust Non-Negative Matrix Factorization for Large-Scale Data Clustering

De Wang, Feiping Nie, and Heng Huang

Presenter: Dr. Heng Huang

Department of Computer Science and Engineering,
The University of Texas at Arlington, TX, USA

July 3, 2016

Table of Contents

1. Motivation
2. New NMF Models
3. Optimization Algorithms
4. Experimental Results

Typical NMF Method

$$\min_{F,G} \left\| X - FG^T \right\|_F^2 \qquad s.t. \quad F \geq 0, G \geq 0 \tag{1.1}$$

where $X \in \mathbb{R}_+^{d \times n}$ is a data matrix with d features and n samples. $F \in \mathbb{R}_+^{d \times c}$ can be viewed as cluster centroids, and $G \in \mathbb{R}_+^{n \times c}$ can be viewed as clustering indicator matrix.

- Converges slowly: hundreds of iterations
- High computational cost: involves large matrix multiplication in each iteration
- Soft clustering: need post processing step to get the final clustering results.
- Not robust to outliers

Table of Contents

1. Motivation
2. New NMF Models
3. Optimization Algorithms
4. Experimental Results

New Fast Robust NMF and NMTF Model

$$RFNMF\_L1 \quad \min_{F\geq 0, G\in Ind} \left\|X - FG^T\right\|_1 \tag{2.1}$$

$$RFNMF \quad \min_{F\geq 0, G\in Ind} \left\|X - FG^T\right\|_{2,1} \tag{2.2}$$

$$RFNMTF \quad \min_{F\in Ind, G\in Ind, S\geq 0} \left\|X - FSG^T\right\|_1 \tag{2.3}$$

where $G \in Ind$ or $F \in Ind$ indicates that G and F are indicator matrices, *i.e.* $g_{ij} = 1$ if x_i belongs to class j, and $g_{ij} = 0$ otherwise.

- Converges fast
- Light computation in each iteration: simple meadian finding plus label assignment
- Hard clustring: no post processing step
- Robust to outliers using $\ell_{2,1}/\ell_1$ loss functions

Laplacian Noise Interpretation for RFNMF_L1

$$x_i = \alpha_i + \sigma_i \tag{2.4}$$

where α_i is the unobservable true data, in NMF clustering it is the clustering centroid, *i.e.* $\alpha_i = FG_{i.}^T$, $G_{i.} \in Ind$. σ_i is the noise. Suppose noise is drawn from Laplacian distribution with zero mean:

$$p(x_i|\alpha_i) = \frac{1}{2b} exp(-\frac{\|x_i - \alpha_i\|_1}{b}) \tag{2.5}$$

where b is the scale parameter of Laplacian distribution.

Interpretation of NMF models

$$\max_{\alpha_i} log \prod_{i=1}^{N} p(x_i|\alpha_i) \Rightarrow \max_{\alpha_i} -\frac{1}{b} \sum_{i=1}^{N} \|x_i - \alpha_i\|_1$$

$$\Rightarrow \quad \min_{\alpha_i} \sum_{i=1}^{N} \|x_i - \alpha_i\|_1 \Rightarrow \min_{F, G_{i.} \in Ind} \sum_{i=1}^{N} \|x_i - FG_{i.}\|_1$$

$$\Rightarrow \quad \min_{F, G \in Ind} \sum_{i=1}^{N} \|X - FG\|_1 \tag{2.6}$$

Similarly:

- RFNMTF can be interpeted from a Laplacian distributed noise, with α replaced by $FSG_{i.}$ instead of $FG_{i.}$
- RFNMF model can be interpreted from a Gaussian distributed noise.

Table of Contents

1 Motivation
2 New NMF Models
3 Optimization Algorithms
4 Experimental Results

Optimization Algorithm for RFNMF_L1

Lemma

Considering the following objective function:

$$\min_{z} \sum_{i} |z - a_i| \tag{3.1}$$

The optimal solution of z is the median value of a_i.

When G fixed:

$$\min_{F \geq 0} \left\| X - FG^T \right\|_1 \tag{3.2}$$

$$\Rightarrow \min_{F \geq 0} \sum_{i} \left\| X_{i.} - \sum_{k} F_{ik} G^T_{.k} \right\|_1$$

$$\Rightarrow \min_{F \geq 0} \sum_{i} \sum_{k} \sum_{G_{jk}=1} |X_{ij} - F_{ik}|$$

Optimization Algorithm for RFNMF_L1

The above problem can be decoupled as: for $\forall i, k$, solving:

$$\min_{F_{ik}} \sum_{G_{jk}=1} |X_{ij} - F_{ik}| \tag{3.3}$$

According to Lemma 1, the optimal solution of F_{iK} can be efficiently obtained by finding the median values of samples belong to the k-th cluster.

When F fixed:

$$g_{ij} = \begin{cases} 1 & j = arg \min_k \|X_{.i} - F_{.k}\|_1 \\ 0 & otherwise \end{cases} \tag{3.4}$$

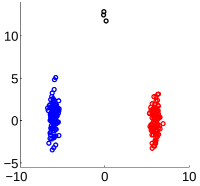
Table of Contents

1. Motivation
2. New NMF Models
3. Optimization Algorithms
4. Experimental Results

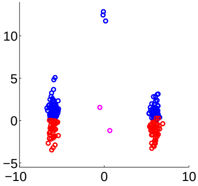
Comparison Methods

(1) Standard NMF (NMF) [4]: solves the objective function in Eq. (1.1).
(2) Orthogonal NMTF (OrthNMF) [2]: factorizes a matrix into three non-negative components, and each column of the soft indicator matrices (F and G) are required to be orthogonal.
(3) SemiNMF [1]: allows the basis matrix F in standard NMF to be mix-signed
(4) Convex NMF (ConvNMF) [1]: restricts the basis matrix F into a linear combination of original data points.
(5) Robust NMF (RNMF) [3]: replaces the loss measurement in standard NMF from Frobenius norm to $\ell_{2,1}$-norm, which makes the model robust to outliers.

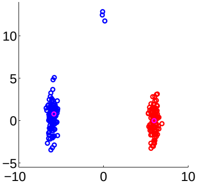
Synthetic Data



(a) Original data with outliers



(b) Clustering results with standard NMF



(c) Clustering results with our three methods

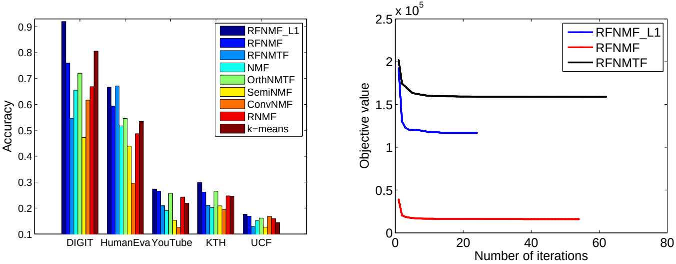
Figure: Clustering performance on synthetic data. Blue points and red points are normal data drawn from two gaussian distributions. Black points are outliers. Magenta points are computed cluster centroids.

Synthetic Data

Table: Average distance from the centroids for normal data (blue and red points in figure 1 (a)), outliers, and all data.

	normal data	outliers	all data
NMF	6.02	10.82	6.09
Our methods	1.27	12.96	1.45

Real World Data Sets



(a) Accuracy (b) DIGIT

Figure: (a): Clustering accuracy (b): objective value versus number of iterations.

Computational Time Comparison

Table: Computational time (in seconds) comparison. Averaged over 10 repetitions.

	RFNMF_L1	RFNMF	RFNMTF	NMF	OrthNMTF	SemiNMF
DIGIT	**3.57**	5.98	12.13	54.89	243.27	46.98
HumanEva	**5.79**	49.32	23.63	67.64	2181.26	12.60
YouTube	**3.70**	**7.86**	20.32	366.91	553.55	43.89
KTH	**8.04**	**12.51**	26.41	300.72	676.98	64.71
UCF	**246.19**	251.20	278.06	1974.57	3891.82	349.75

- C. Ding, T. Li, and M. I. Jordan.
  Convex and semi-nonnegative matrix factorizations.
  Pattern Analysis and Machine Intelligence, IEEE Transactions on, 32(1):45–55, 2010.

- C. Ding, T. Li, W. Peng, and H. Park.
  Orthogonal nonnegative matrix t-factorizations for clustering.
  In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 126–135. ACM, 2006.

- D. Kong, C. Ding, and H. Huang.
  Robust nonnegative matrix factorization using l2,1-norm.
  In *ACM Conference on Information and Knowledge Management (CIKM 2011)*, 2011.

- D. Lee and H. Seung.
  Algorithms for nonnegative matrix factorization.
  In *NIPS*, pages 556–562, 2001.