Graph Regularized Non-negative Matrix Factorization for Data Representation

Authors: D. Cai, X. He, J.Han and Thomas S. Huang,2011

Reporter:Yunfei Wang

Sep.21,2012

Outline

- NMF with Manifold Regularization
- Updating Rules
- Computational Complexity Analysis
- Experimental results
- Proofs of convergence
- Weighted NMF and GNMF

NMF with Manifold Regularization

- Local invariance assumption
  - Two data points close in the intrinsic geometry of the data distribution, the respect to the new basis are also close to each other.
- Drawback of NMF
  - Failing to discover the intrinsic geometrical and discriminating structure of the data space
- GNMF
  - Incorporating a geometrically based regularizer into NMF.

Nearest Neighbor Graph

- Model the geometric structure effectively
- Weight Matrix

  1) 0-1 weighting $$W_{ij} = \begin{cases} 1, \text{nodes i and j are close} \\ 0, otherwise \end{cases}$$

  2) Heat kernel weighting $$\boldsymbol{W}_{ij} = \boldsymbol{e}^{-\frac{\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2}{\sigma}}$$

  3) Dot-product weighting $$\boldsymbol{W}_{ij} = \boldsymbol{x}_j^T \boldsymbol{x}_i$$

Measure the dissimilarity in NMF

- Euclidean distance

$$d(\mathbf{z}_j, \mathbf{z}_l) = \|\mathbf{z}_j - \mathbf{z}_l\|^2$$

- Divergence

$$D(\mathbf{z}_j \| \mathbf{z}_l) = \sum_{k=1}^{K} \left(v_{jk} \log \frac{v_{jk}}{v_{lk}} - v_{jk} + v_{lk} \right)$$

Measure the smoothness

- Euclidean distance

$$\mathcal{R}_1 = \frac{1}{2}\sum_{j,l=1}^{N} \|\mathbf{z}_j - \mathbf{z}_l\|^2 \mathbf{W}_{jl}$$

$$= \sum_{j=1}^{N} \mathbf{z}_j^T \mathbf{z}_j \mathbf{D}_{jj} - \sum_{j,l=1}^{N} \mathbf{z}_j^T \mathbf{z}_l \mathbf{W}_{jl}$$

$$= \mathrm{Tr}(\mathbf{V}^T \mathbf{D} \mathbf{V}) - \mathrm{Tr}(\mathbf{V}^T \mathbf{W} \mathbf{V}) = \mathrm{Tr}(\mathbf{V}^T \mathbf{L} \mathbf{V})$$

- Divergence

$$\mathcal{R}_2 = \frac{1}{2}\sum_{j,l=1}^{N} \Big(D(\mathbf{z}_j \| \mathbf{z}_l) + D(\mathbf{z}_l \| \mathbf{z}_j) \Big) \mathbf{W}_{jl}$$

$$= \frac{1}{2}\sum_{j,l=1}^{N}\sum_{k=1}^{K} \left(v_{jk} \log \frac{v_{jk}}{v_{lk}} + v_{lk} \log \frac{v_{lk}}{v_{jk}} \right) \mathbf{W}_{jl}$$

Objective functions in GNMF

- Euclidean distance

$$\begin{aligned}\mathcal{O}_1 &= \operatorname{Tr}\left((\mathbf{X}-\mathbf{U}\mathbf{V}^T)(\mathbf{X}-\mathbf{U}\mathbf{V}^T)^T\right) + \lambda \operatorname{Tr}(\mathbf{V}^T\mathbf{L}\mathbf{V}) \\ &= \operatorname{Tr}\left(\mathbf{X}\mathbf{X}^T\right) - 2\operatorname{Tr}\left(\mathbf{X}\mathbf{V}\mathbf{U}^T\right) + \operatorname{Tr}\left(\mathbf{U}\mathbf{V}^T\mathbf{V}\mathbf{U}^T\right) \\ &\quad + \lambda \operatorname{Tr}(\mathbf{V}^T\mathbf{L}\mathbf{V})\end{aligned}$$

- Divergence

$$\begin{aligned}\mathcal{O}_2 = &\sum_{i=1}^{M}\sum_{j=1}^{N}\left(x_{ij}\log\frac{x_{ij}}{\sum_{k=1}^{K}u_{ik}v_{jk}} - x_{ij} + \sum_{k=1}^{K}u_{ik}v_{jk}\right) \\ &+ \frac{\lambda}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{k=1}^{K}\left(v_{jk}\log\frac{v_{jk}}{v_{lk}} + v_{lk}\log\frac{v_{lk}}{v_{jk}}\right)\mathbf{W}_{jl}\end{aligned}$$

Updating Rules

- Euclidean distance

$$u_{ik} \leftarrow u_{ik} \frac{(\mathbf{XV})_{ik}}{(\mathbf{UV}^T\mathbf{V})_{ik}} \qquad v_{jk} \leftarrow v_{jk} \frac{(\mathbf{X}^T\mathbf{U} + \lambda \mathbf{WV})_{jk}}{(\mathbf{VU}^T\mathbf{U} + \lambda \mathbf{DV})_{jk}}$$

- Divergence

$$u_{ik} \leftarrow u_{ik} \frac{\sum_j \left(x_{ij} v_{jk} / \sum_k u_{ik} v_{jk}\right)}{\sum_j v_{jk}}$$

$$\mathbf{v}_k \leftarrow \left(\sum_i u_{ik}\mathbf{I} + \lambda \mathbf{L}\right)^{-1} \begin{bmatrix} v_{1k} \sum_i \left(x_{i1} u_{ik} / \sum_k u_{ik} v_{1k}\right) \\ v_{2k} \sum_i \left(x_{i2} u_{ik} / \sum_k u_{ik} v_{2k}\right) \\ \vdots \\ v_{Nk} \sum_i \left(x_{iN} u_{ik} / \sum_k u_{ik} v_{Nk}\right) \end{bmatrix}$$

Computational Complexity Analysis

Computational operation counts for each iteration in NMF and GNMF

	F-norm formulation			
	fladd	flmlt	fldiv	overall
NMF	$2MNK + 2(M+N)K^2$	$2MNK + 2(M+N)K^2 + (M+N)K$	$(M+N)K$	$O(MNK)$
GNMF	$2MNK + 2(M+N)K^2$ $+N(p+3)K$	$2MNK + 2(M+N)K^2 + (M+N)K$ $+N(p+1)K$	$(M+N)K$	$O(MNK)$

	Divergence formulation			
	fladd	flmlt	fldiv	overall
NMF	$4MNK + (M+N)K$	$4MNK + (M+N)K$	$2MN + (M+N)K$	$O(MNK)$
GNMF	$4MNK + (M+2N)K$ $+q(p+4)NK$	$4MNK + (M+N)K$ $+Np + q(p+4)NK$	$2MN + MK$	$O\big((M+q(p+4))NK\big)$

fladd: a floating-point addition — flmlt: a floating-point multiplication — fldiv: a floating-point division

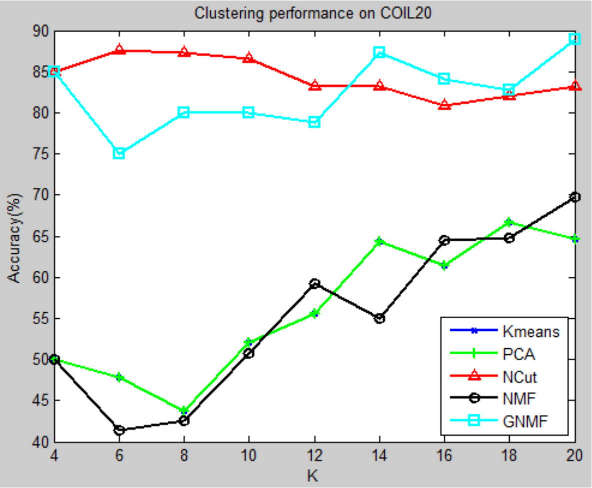
N: the number of sample points — M: the number of features — K: the number of factors

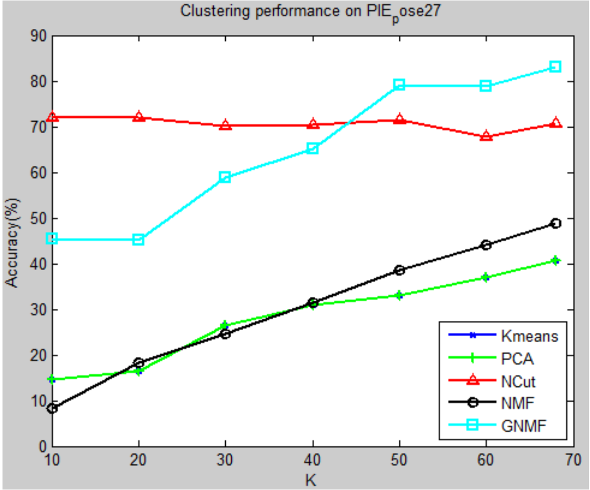
p: the number of nearest neighbors — q: the number of iterations in CG

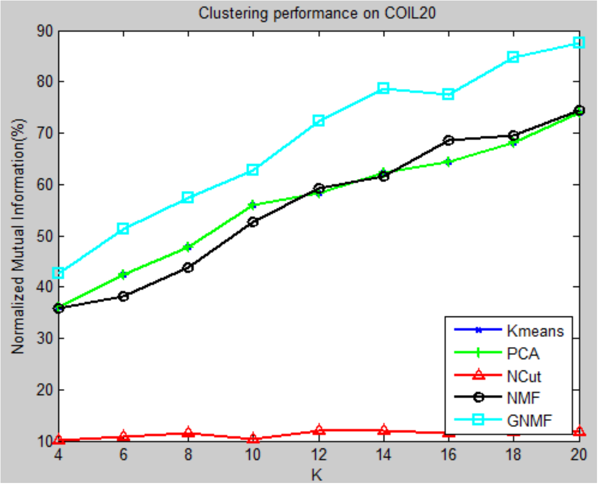
Proofs of convergence

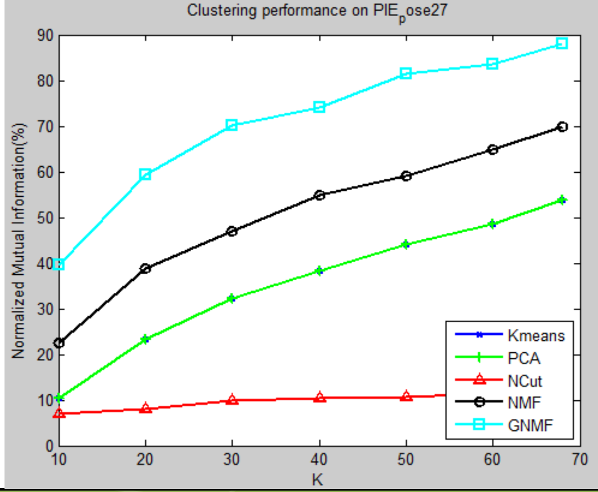
- The proving steps are similar to NMF
  - Define an auxiliary function
  - Construct an auxiliary function for objective function
  - Prove the function constructed above satisfy all the conditions.
  - Get the update rules by setting the gradient of auxiliary function to zero.

Experimental results

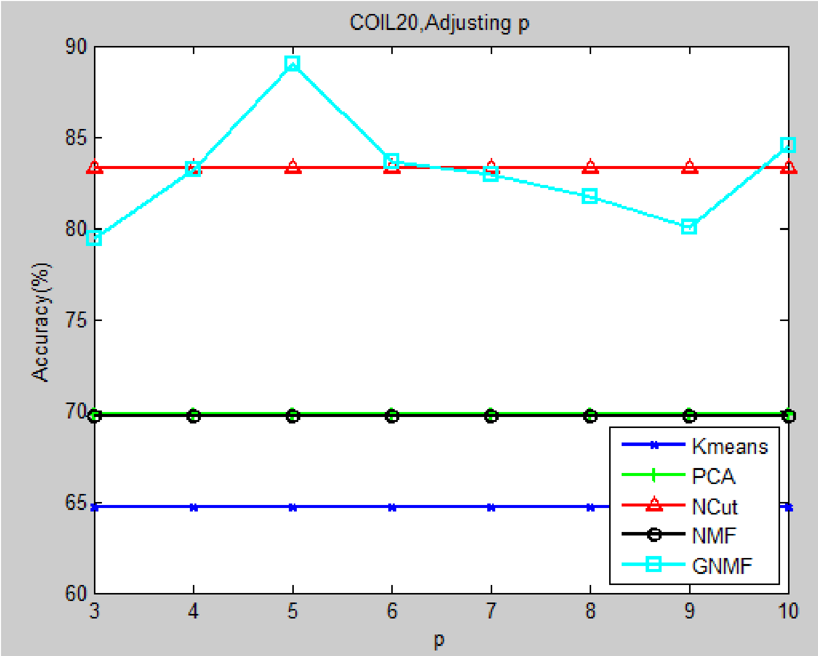


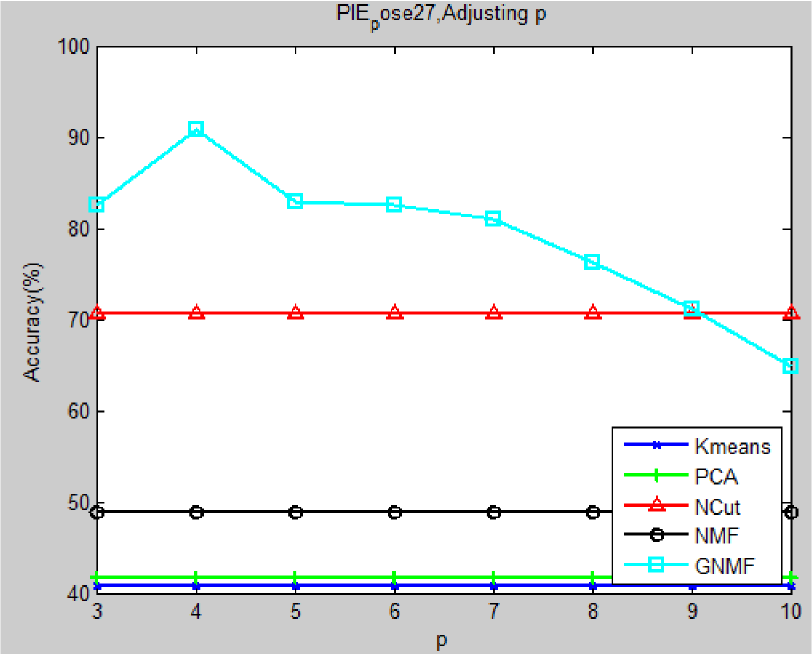




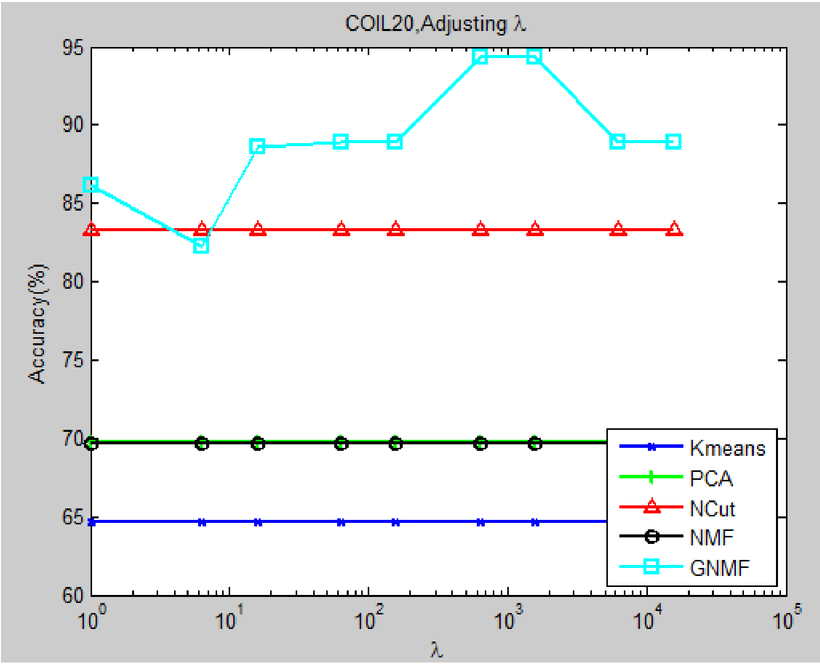


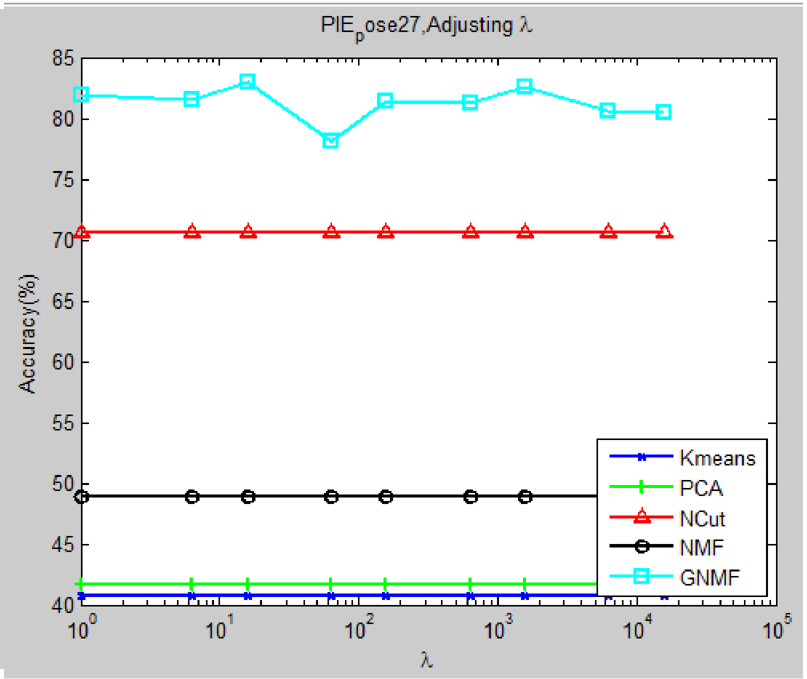
Experimental results



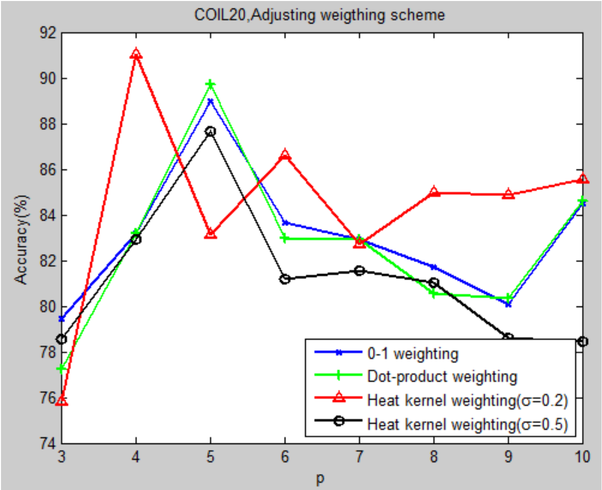


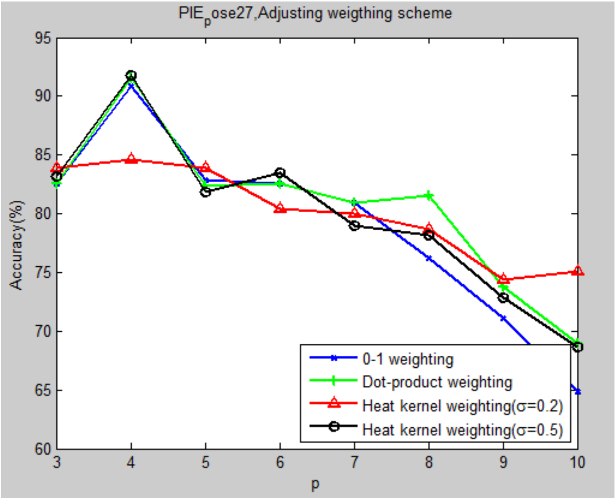
Experimental results

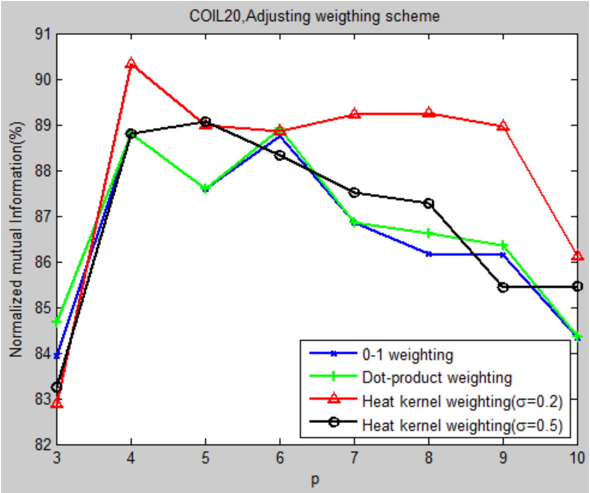


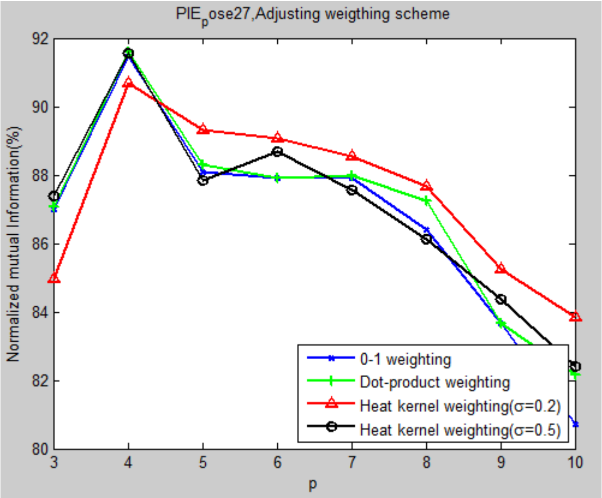


Experimental results









Weighted NMF and GNMF

- Weighted NMF

$$
\begin{aligned}
O' &= \sum_{j=1}^{N} \gamma_j (\mathbf{x}_j - \mathbf{U}\mathbf{z}_j)^T (\mathbf{x}_j - \mathbf{U}\mathbf{z}_j) \\
&= \mathrm{Tr}\left((\mathbf{X} - \mathbf{U}\mathbf{V}^T) \Gamma (\mathbf{X} - \mathbf{U}\mathbf{V}^T)^T \right) \\
&= \mathrm{Tr}\left((\mathbf{X}\Gamma^{1/2} - \mathbf{U}\mathbf{V}^T\Gamma^{1/2}) (\mathbf{X}\Gamma^{1/2} - \mathbf{U}\mathbf{V}^T\Gamma^{1/2})^T \right) \\
&= \mathrm{Tr}\left((\mathbf{X}' - \mathbf{U}\mathbf{V}'^T)^T (\mathbf{X}' - \mathbf{U}\mathbf{V}'^T) \right)
\end{aligned}
$$

- Weighted GNMF

$$
\begin{aligned}
\mathcal{O}' &= \sum_{j=1}^{N} \gamma_j (\mathbf{x}_j - \mathbf{U}\mathbf{z}_j)^T (\mathbf{x}_j - \mathbf{U}\mathbf{z}_j) + \lambda \,\mathrm{Tr}(\mathbf{V}^T\mathbf{L}\mathbf{V}) \\
&= \mathrm{Tr}\left((\mathbf{X} - \mathbf{U}\mathbf{V}^T) \Gamma (\mathbf{X} - \mathbf{U}\mathbf{V}^T)^T \right) + \lambda \,\mathrm{Tr}(\mathbf{V}^T\mathbf{L}\mathbf{V}) \\
&= \mathrm{Tr}\left((\mathbf{X}\Gamma^{1/2} - \mathbf{U}\mathbf{V}^T\Gamma^{1/2}) (\mathbf{X}\Gamma^{1/2} - \mathbf{U}\mathbf{V}^T\Gamma^{1/2})^T \right) \\
&\quad + \lambda \,\mathrm{Tr}(\mathbf{V}^T\mathbf{L}\mathbf{V}) \\
&= \mathrm{Tr}\left((\mathbf{X}' - \mathbf{U}\mathbf{V}'^T)^T (\mathbf{X}' - \mathbf{U}\mathbf{V}'^T) \right) + \lambda \,\mathrm{Tr}(\mathbf{V}'^T\mathbf{L}'\mathbf{V}')
\end{aligned}
$$