Efficient Rank-one Residue Approximation Method for Graph Regularized Non-negative Matrix Factorization

PKDD'2013

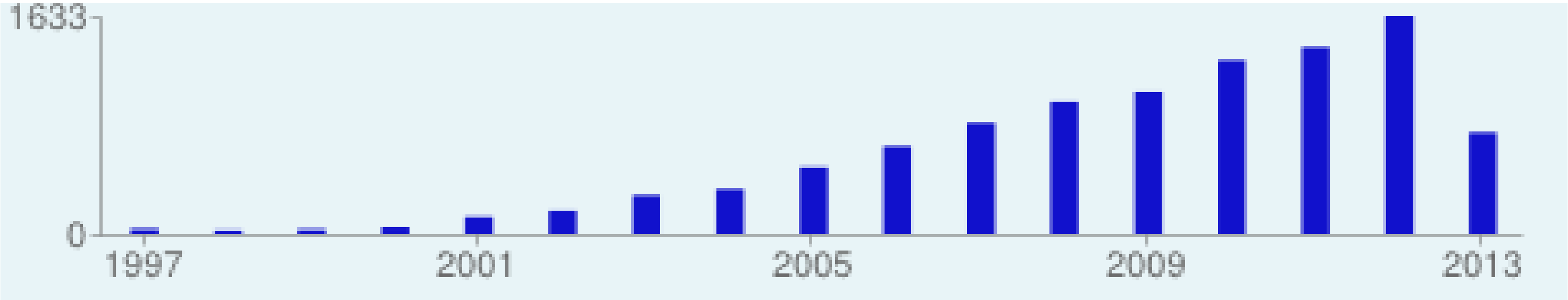
Presenter: Qing LIAO

Dept. of Computer Science and Engineering, The Hong Kong University of Science and Technology

Contents

Background

Research hotspot



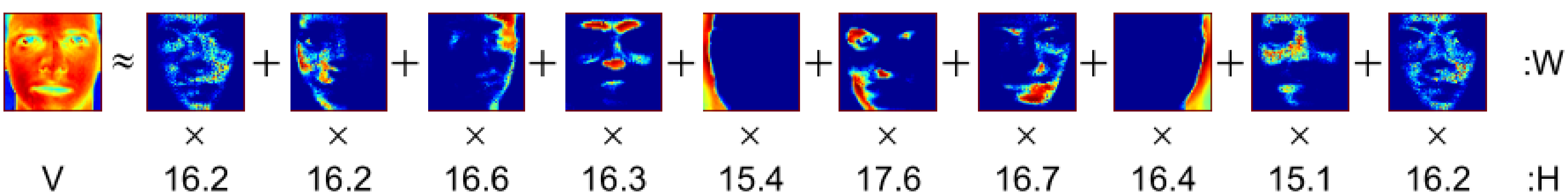
Reference trend statistics (Lee & Seung, Nature, 1999) until July 2013

Background

Nonnegative matrix factorization (NMF)

$$\boldsymbol{V} \geq \mathbf{0} \approx \boldsymbol{W} \geq \mathbf{0} \times \boldsymbol{H} \geq \mathbf{0}$$

Two Advantages: (a) Dimension Reduction
(b) Parts-based Representation



Background

Applications

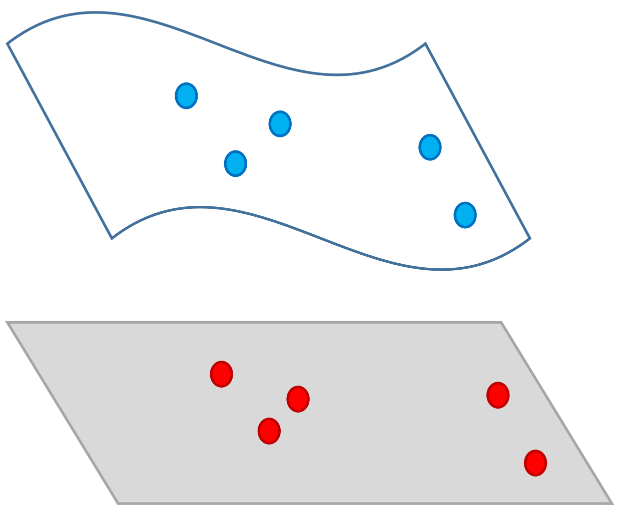
- Pattern recognition
- Data mining
- Image annotation
- video monitoring
- speech analysis
- High spectral analysis
- music analysis

Background

Drawbacks of NMF

- Ignore the property of the data set
- Lose the geometric structure of the data set

Graph regularized NMF (GNMF)



Background

Multiplicative update rule (MUR)

$$U_{t+1} = U_t \circ \frac{XV_t}{U_t V_t^T V_t}$$

$$V_{t+1} = V_t \circ \frac{X^T U_{t+1} + \beta W V_t}{V_t U_{t+1}^T U_{t+1} + \beta D V_t}$$

Drawbacks of MUR

- Slow convergence
- Non-stationarity point

Background

State-of-the-art NeNMF

- Tolerance is too small, large computational cost
- Tolerance is too large, low-quality solution

$$H^{t+1} = \arg\min_{H \geq 0} F(W^t, H) = \frac{1}{2}\|X - W^t H\|_F^2$$

$$W^{t+1} = \arg\min_{W \geq 0} F(W, H^{t+1}) = \frac{1}{2}\left\|X^T - {H^{t+1}}^T W^T\right\|_F^2$$

Background

Rank-one residue approximation (RRA)

More Efficient

RRA updates each of matrix column by approximating the residue matrix by their outer product

Stationary Point

We theoretically prove our method can converge rapidly to a stationary point

Graph regularized NMF (GNMF)

GNMF

The objective function of NMF

$$\min_{U\geq 0, V\geq 0} \frac{1}{2}\|X - UV^T\|_F^2$$

Preserve the geometrical structure

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\|v_i - v_j\|_2^2 W_{ij} = tr(V^T L V)$$

The objective function of GNMF

$$\min_{U\geq 0, V\geq 0} f(U,V) = \frac{1}{2}\|X - UV^T\|_F^2 + \frac{\beta}{2} tr(V^T L V)$$

Rank-one Residue Approximation for GNMF

Rank-one Residue Approximation (RRA)

- Two sub-problems

$$\min_{U_{\cdot k} \geq 0} \frac{1}{2} \|R_k - U_{\cdot k} V_{\cdot k}^T\|_F^2 \quad (1)$$

$$R_k = X - \sum_{l \neq k} U_{.l} V_{.l}^T$$

$$Tr(V^T L V) = \sum_{k=1}^{r} V_{.k}^T L V_{.k}$$

$$\min_{V_{\cdot k} \geq 0} \frac{1}{2} \|R_k - U_{\cdot k} V_{\cdot k}^T\|_F^2 + \frac{\beta}{2} V_{\cdot k}^T L V_{\cdot k} \quad (2)$$

- Sub-problem (1) has a closed-form solution(Cai,2008)

$$U_{\cdot k} = \frac{\prod_{+}(R_k V_{\cdot k})}{\|V_{\cdot k}\|_2^2}$$

- Sub-problem (2) applies the Lagrangian multiplier method

$$V_{\cdot k} = \prod_{+}((\|U_{\cdot k}\|_2^2 I + \beta L)^{-1} R_k^T U_{\cdot k})$$

Rank-one Residue Approximation for GNMF

Stationarity point

- **Proposition 1** *Every limited point generated by alternating (1) and (2) is a stationary point.*
- ***Proof***

The GNMF problem can be written as a bound-constrained optimization problem according to (Lin, 2007)

$$\min_{[U,V]\in\Omega} \frac{1}{2}\|X - \sum_{k=1}^{r} U_{\cdot k}V_{\cdot k}^T\|_F^2 + \frac{\beta}{2}\sum_{k=1}^{r} V_{\cdot k}^T L V_{\cdot k}$$

Where $\Omega = \prod_{k=1}^{r} \Omega_k^U \times \prod_{k=1}^{r} \Omega_k^V$ is a Cartesian product of

Proposition 2.7.1 in (Bertsekas, 1999): Suppose that f is continuously differentiable over the set X of Eq.(2.107). Furthermore, suppose that for each i and $x \in X$, the minimum below

$$min_{\varepsilon \in x_i} f(x_1, \dots, x_{i-1}, \varepsilon, x_{i+1}, \dots, x_m)$$

is uniquely attained, Let $\{x^k\}$ be the sequence generated by the block coordinate descent method (2.108). Then, every limit point of $\{x^k\}$ is a stationary point

Rank-one Residue Approximation for GNMF

Algorithm 1 Rank-one Residue Approximation for GNMF

Input: $X \in \mathbb{R}_{+}^{m \times n}, L \in \mathbb{R}^{n \times n}, 1 \leq r \leq \min(m, n), \beta$
Output: $U \in \mathbb{R}_{+}^{m \times r}, V \in \mathbb{R}_{+}^{n \times r}$

Initialize: $U^1 \in R_{+}^{m \times r}, V^1 \in R_{+}^{n \times r}, t = 1$
Calculate: $L = \Theta \Sigma \Theta^T \approx \tilde{\Theta} \tilde{\Sigma} \tilde{\Theta}^T, R^1 = X - U^1 {V^1}^T$
repeat
  for $k = 1 \ldots r$ **do**
    Calculate: $R_k^t = R^t + U_{\cdot k}^t {V_{\cdot k}^t}^T$
    Update: $U_{\cdot k}^{t+1} = \prod_{+}(R_k^t V_{\cdot k}^t))/\|V_{\cdot k}^t\|_2^2)$
    Calculate: $A_k^t \approx (\|U_{\cdot k}^{t+1}\|_2^2 I + \beta L)^{-1}$ ← Sherman-Morrison-Woodbury(SMW)
    Update: $V_{\cdot k}^{t+1} = \prod_{+}(A_k^t {R_k^t}^T U_{\cdot k}^{t+1})$
    Update: $R^t = R_k^t - U_{\cdot k}^{t+1} {V_{\cdot k}^{t+1}}^T$
  end for
  Update: $R^{t+1} = R^t$
  Update: $t \leftarrow t + 1$
until The Stopping Condition (18) is Satisfied.
$U = U^t, V = V^t$

Evaluation

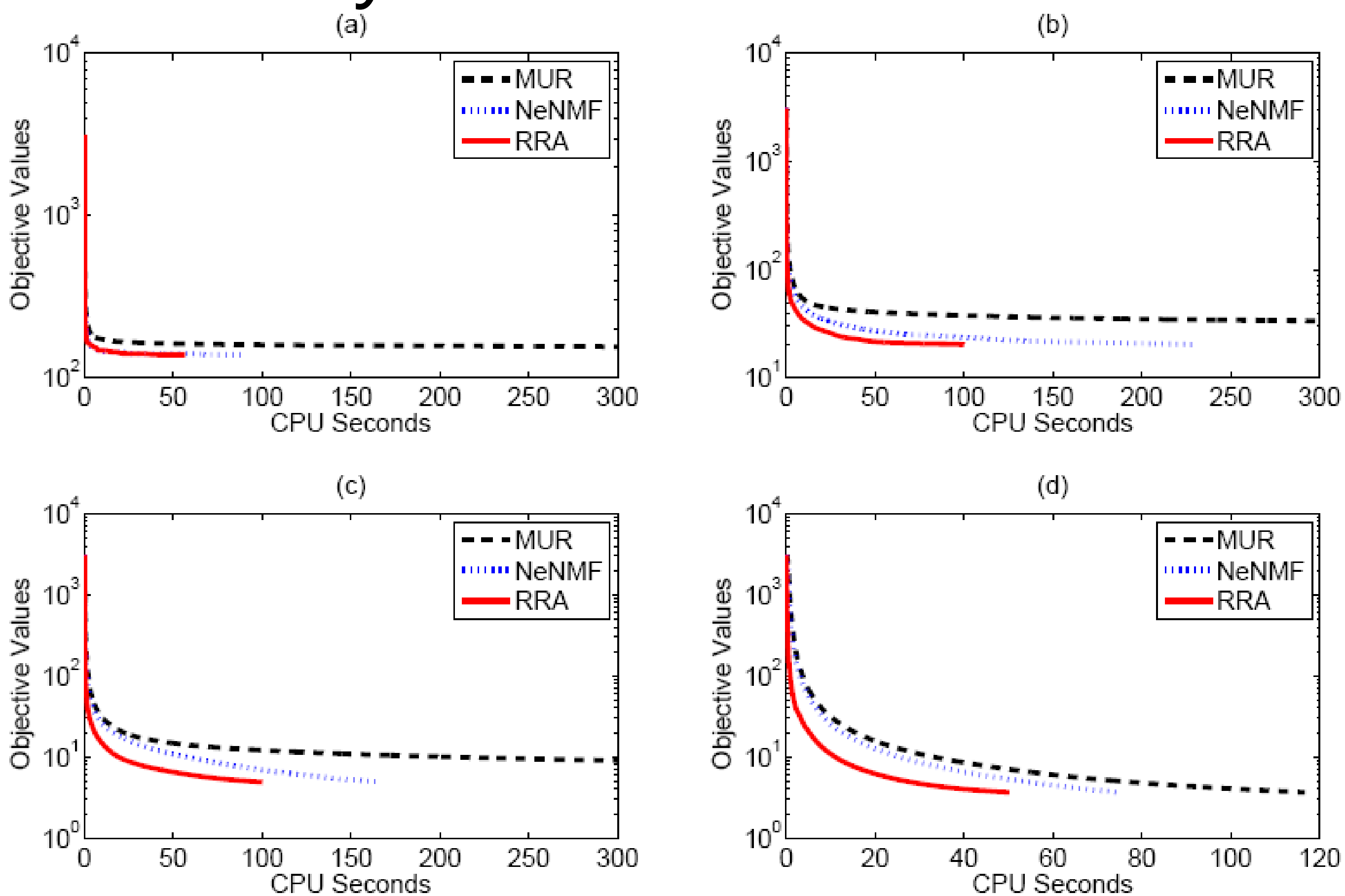
Data sets

- COIL-20 image library
  - COIL-20 contains images of 20 objects viewed from different angles
  - Totally 72 images were taken for each object
- CMU PIE2 face image database
  - CMU PIE2 contains face images of 68 individuals
  - Totally 42 facial images for each individual taken under different lighting and illumination conditions
- COIL-20 composes of a 1024×1440 matrix and CMU PIE composes of a 1024×2856 matrix

COIL-20: http://www1.cs.columbia.edu/CAVE/software/softlib/coil-20.php
CMU PIE2: http://www.ri.cmu.edu/projects/project 418.html

Evaluation

Efficiency Evaluation

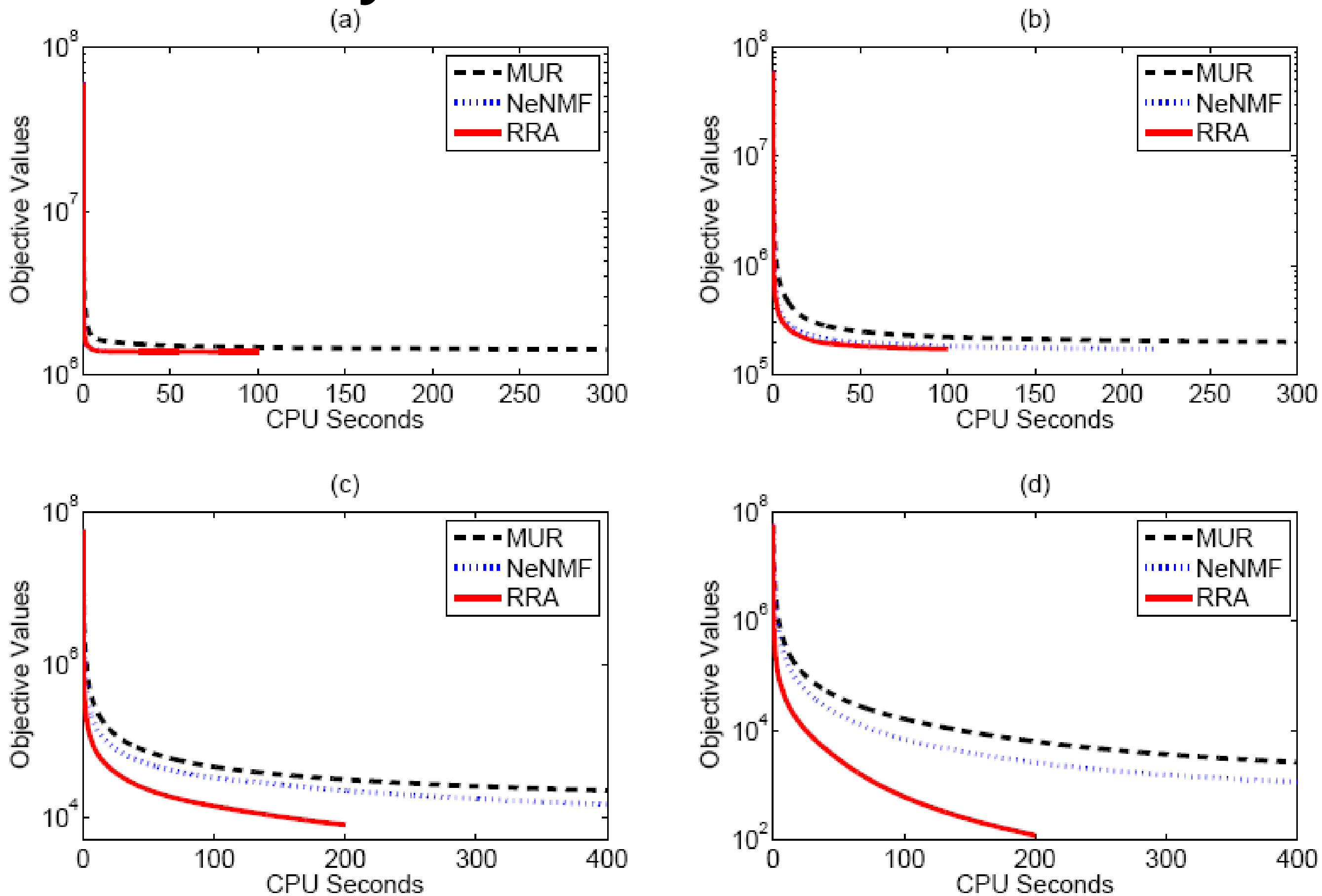


K=5
(a) r=20
(b) r=50
(c) r=100
(d) r=200

RRA performs much more rapidly than MUR and NeNMF

Evaluation

Efficiency Evaluation

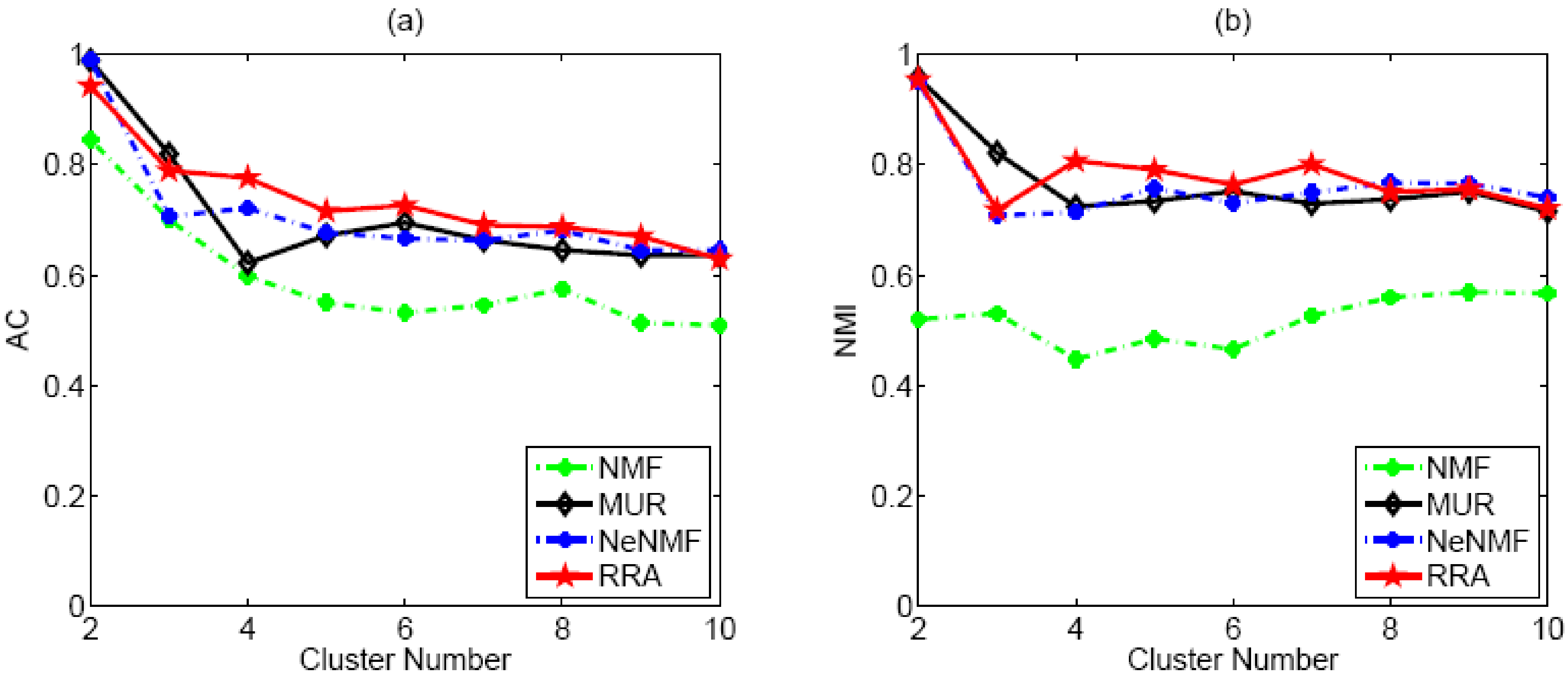


K=5
(a) r=20
(b) r=50
(c) r=100
(d) r=200

RRA performs much more rapidly than MUR and NeNMF

Evaluation

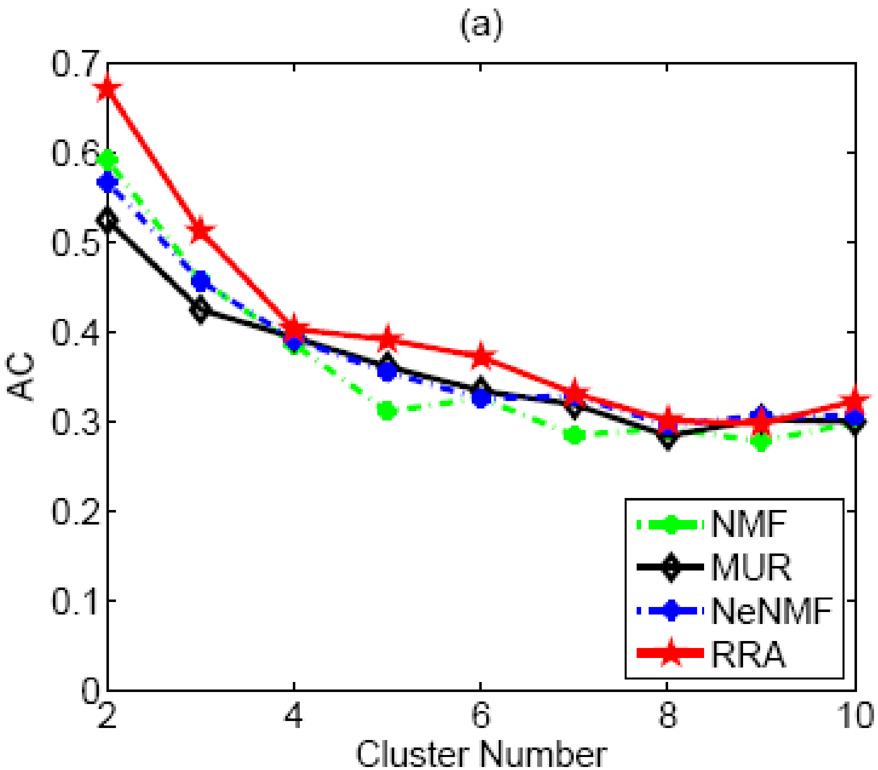
Image Clustering

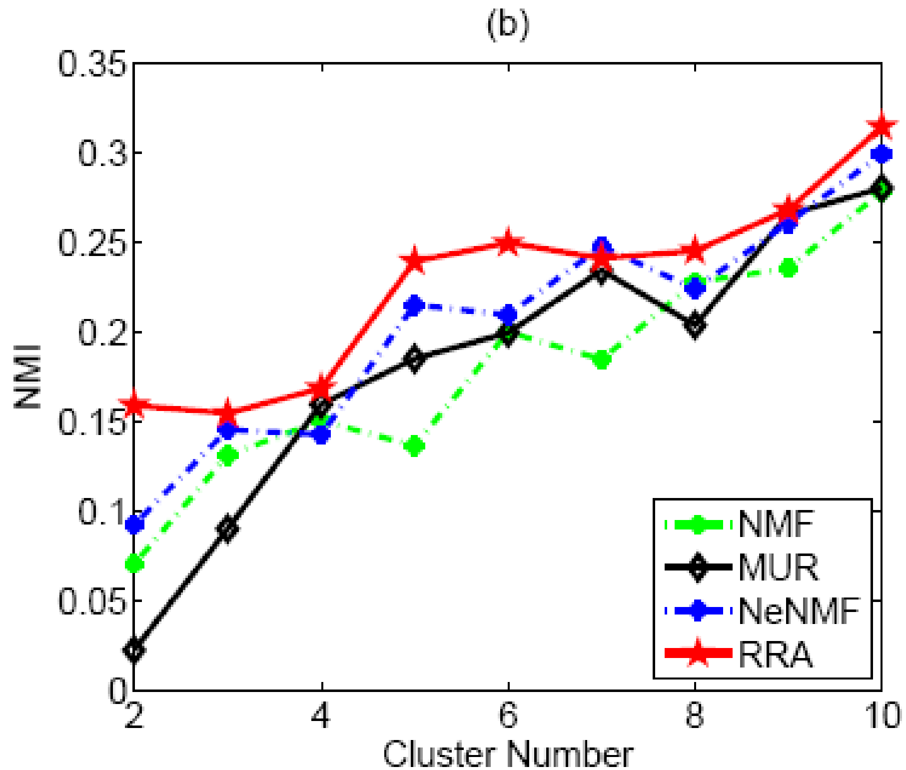


RRA slightly outperforms MUR and NeNMF in terms of AC and NMI

Evaluation

Image Clustering





RRA slightly outperforms MUR and NeNMF in terms of AC and NMI

Conclusion

We propose a novel efficient rank-one residue approximation (RRA) solver for graph regularized non-negative matrix factorization (GNMF)

- Faster
  - Unlike the existing GNMF solvers which recursively update each factor matrix as a whole, RRA recursively updates each column of both factor matrices in an analytic formulation
- RRA theoretically converges to a stationary point of GNMF

Future Work

Community discovery

- Our method is suitable for community discovery, because the cluster performance is quite good from our experiments

Extend to Large Scale Dataset

- Memory complexity is our future work
- Laplacian matrix is a dense matrix so that we cannot store it in sparsity

Q&A
Thanks!

Speed Rate V.S. Scale

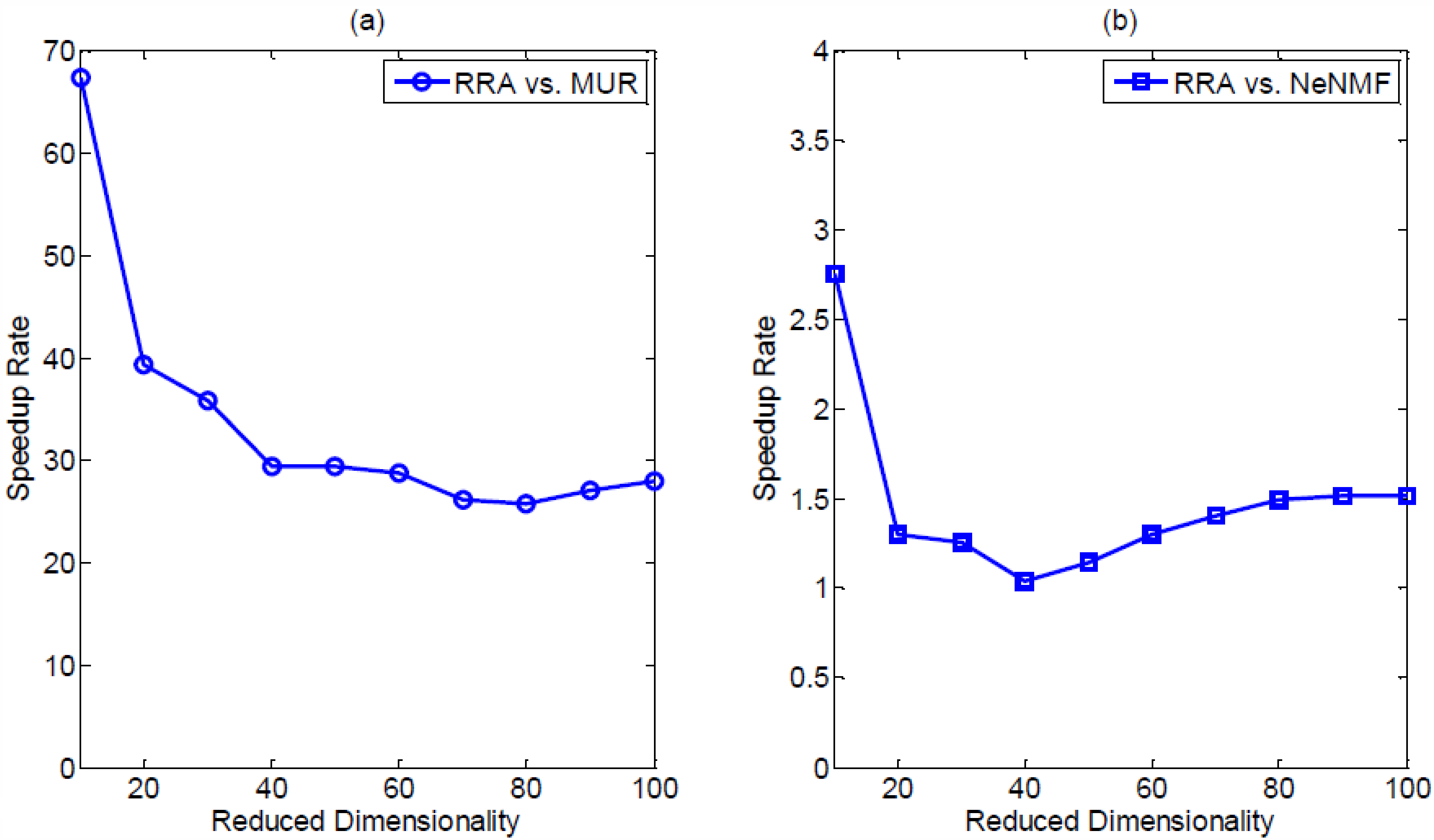


Fig. 5. Speedup rate versus reduced dimensionality: (a) RRA versus MUR and (b) RRA versus NeNMF