电子科技大学
University of Electronic Science and Technology of China

# Paper Sharing

## L-EnsNMF: Boosted Local Topic Discovery via Ensemble of Nonnegative Matrix Factorization

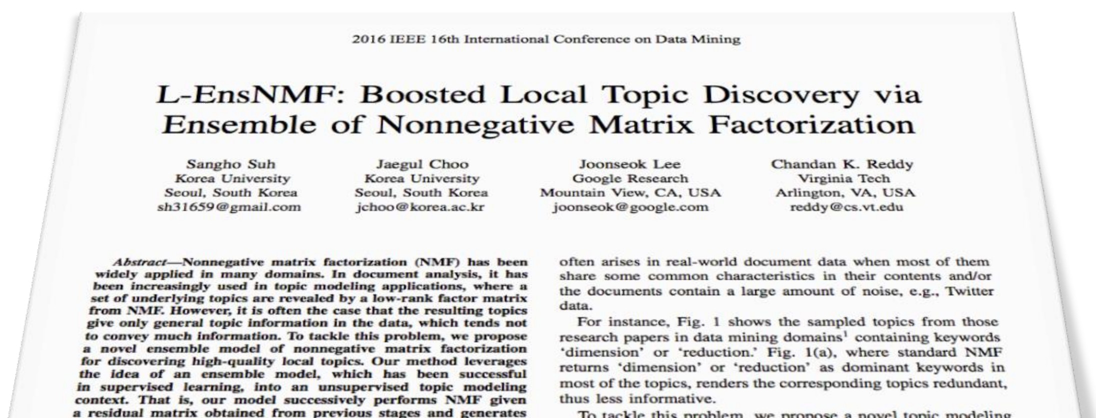
*College of Computer Science*
高崇铭（*Chongming Gao*）

# Background of L-EnsNMF

Sangho Suh, Korea University

*09/01 2017* Joining CrowdLab at Univ. of Waterloo as PhD student in Computer Science

- *09/07 2016* "L-EnsNMF" Paper accepted to ICDM'16 (8.5% acceptance rate)!
- *12/14 2016* Received Best Student Paper Award for "L-EnsNMF" at ICDM'16 Banquet
- *12/26 2016* Accepted invitation to submit an abridged version of "L-EnsNMF" to Sister Conference Best Paper Track at IJCAI'17 (Melbourn, Australia)
- *04/28 2017* "L-EnsNMF" paper accepted to Sister Conference Best Paper Track at IJCAI'17



2016 IEEE 16th International Conference on Data Mining

L-EnsNMF: Boosted Local Topic Discovery via Ensemble of Nonnegative Matrix Factorization

Sangho Suh, Korea University, Seoul, South Korea, sh31659@gmail.com
Jaegul Choo, Korea University, Seoul, South Korea, jchoo@korea.ac.kr
Joonseok Lee, Google Research, Mountain View, CA, USA, joonseok@google.com
Chandan K. Reddy, Virginia Tech, Arlington, VA, USA, reddy@cs.vt.edu

*Abstract*—Nonnegative matrix factorization (NMF) has been widely applied in many domains. In document analysis, it has been increasingly used in topic modeling applications, where a set of underlying topics are revealed by a low-rank factor matrix from NMF. However, it is often the case that the resulting topics give only general topic information in the data, which tends not to convey much information. To tackle this problem, we propose a novel ensemble model of nonnegative matrix factorization for discovering high-quality local topics. Our method leverages the idea of an ensemble model, which has been successful in supervised learning, into an unsupervised topic modeling context. That is, our model successively performs NMF given a residual matrix obtained from previous stages and generates

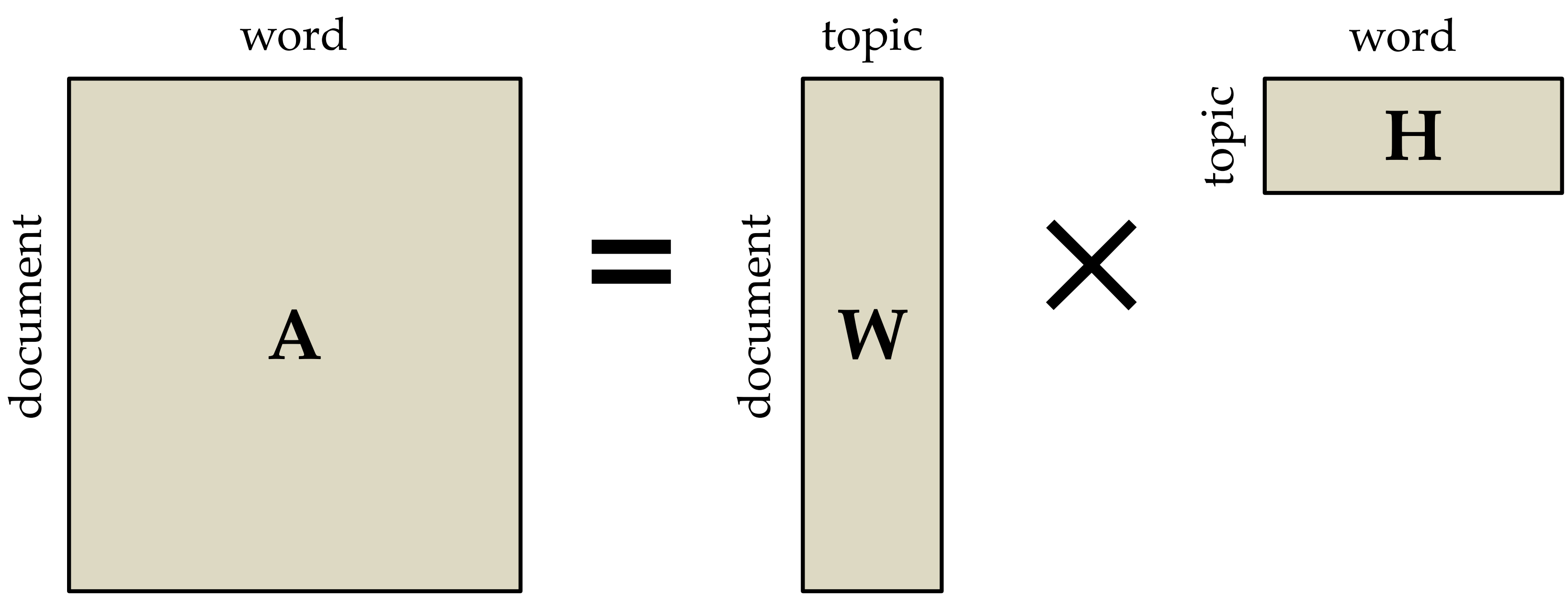
often arises in real-world document data when most of them share some common characteristics in their contents and/or the documents contain a large amount of noise, e.g., Twitter data.

For instance, Fig. 1 shows the sampled topics from those research papers in data mining domains[1] containing keywords 'dimension' or 'reduction.' Fig. 1(a), where standard NMF returns 'dimension' or 'reduction' as dominant keywords in most of the topics, renders the corresponding topics redundant, thus less informative.

# 1. Motivation

Nonnegative matrix factorization (NMF) has been widely applied in many domains. In document analysis, it has been increasingly used in topic modeling applications, where a set of underlying topics are revealed by a low-rank factor matrix from NMF.

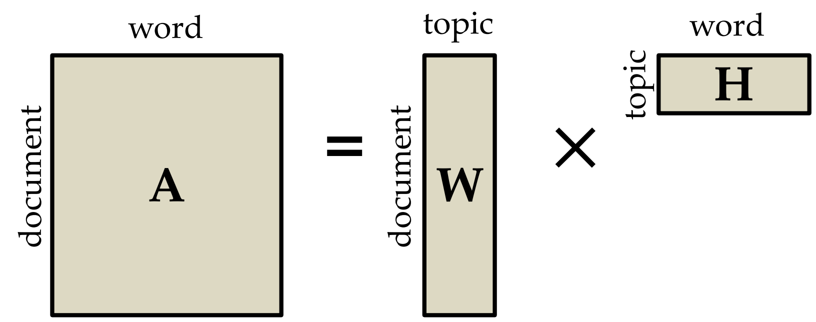
$$\min_{W,H\geq 0} \|X - WH\|_F^2 .$$



# 1. Motivation

Question

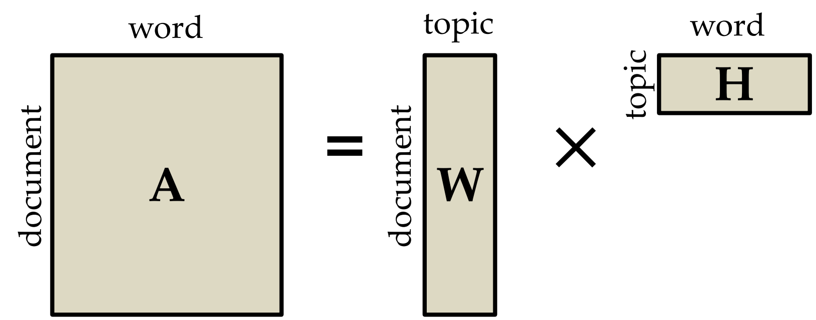
Why NMF became so popular?

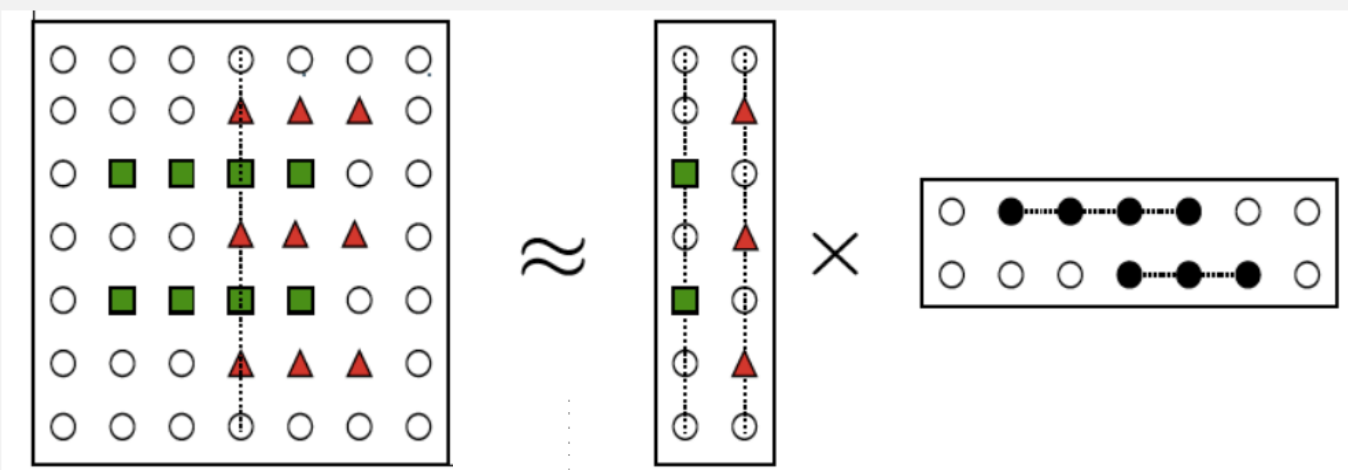


# 1. Motivation

**Question**

Why NMF became so popular?



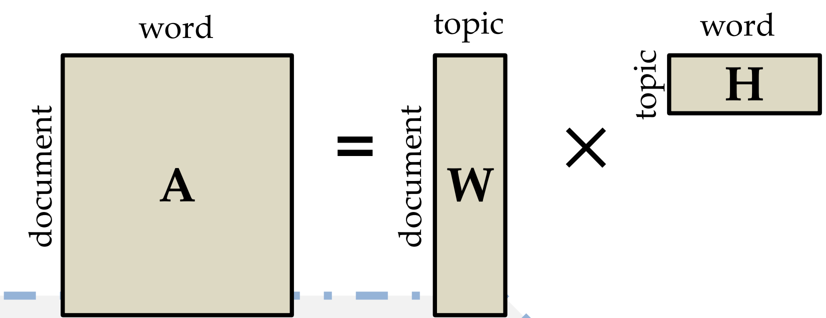



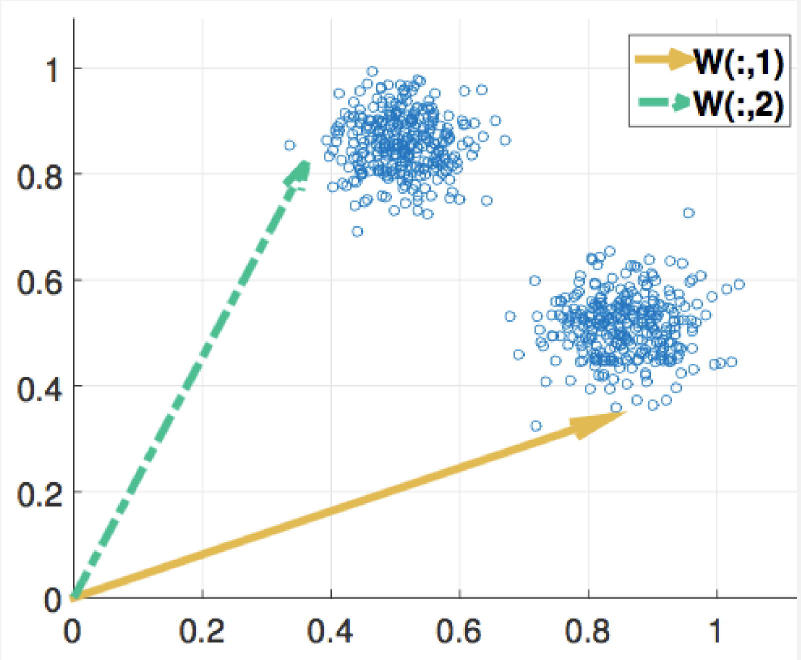
For the strong capability of explanation!

# 1. Motivation

**Question**

Why NMF became so popular?




For the strong capability of explanation!

# 1. Motivation

## Limitations of Traditional NMF

- It is often the case that the resulting topics give only general topic information in the data, which tends not to convey much information.

- Such results may give dominant topics that are highly redundant with each other.

## Contributions of *L-EnsNMF*

- It develops an ensemble approach of nonnegative matrix factorization based on a gradient-boosting framework. This novel approach can extract high-quality local topics from noisy documents dominated by a few uninteresting topics.

# 1. Motivation

## Toy Example:

Fig. 1 shows the sampled topics from those research papers in data mining domains1 containing keywords ‘dimension’ or ‘reduction.’ Fig. 1(a), where standard NMF returns ‘dimension’ or ‘reduction’ as dominant keywords in most of the topics, renders the corresponding topics redundant, thus less informative.

By contrast, L-EnsNMF can reveal not only dominant topics but also minor but meaningful, important topics.

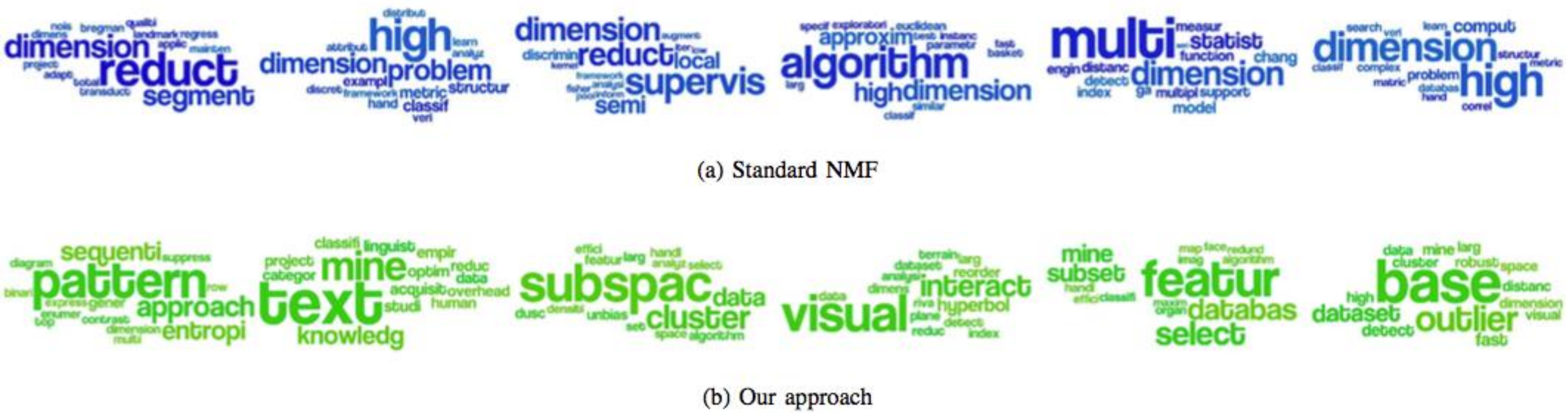



Fig. 1: Topic examples extracted from research papers in the data mining area published in 2000 - 2008

# 2. Method

## Gradient Boosting

**Core Principle**：

In the iterative process, the RESIDUE of last fitting will be the objective value of this fitting step

Objective value in step *m*

Current model output

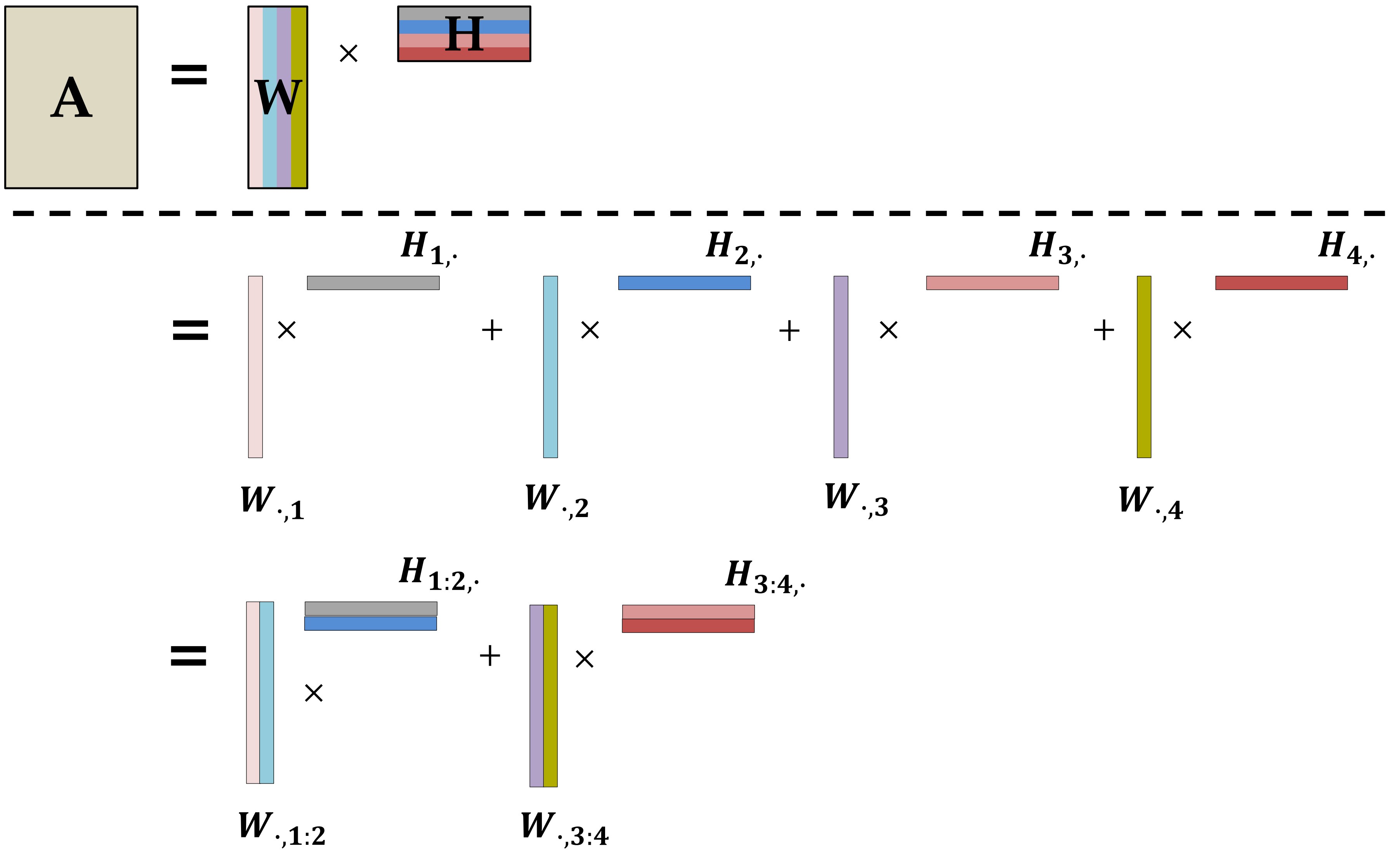
Objective value in step *m+1*

$$F_{m+1}(x) = F_m(x) + h(x) = y$$

Ground truth value

# 2. Method

## One important property of matrix multiplication



# 2. Method

电子科技大学

Objective function of traditional NMF

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

$$\min_{W^{(i)},H^{(i)}\geq 0,\, i=1,\cdots,q} \left\| X - \sum_{i=1}^{q} W^{(i)} H^{(i)} \right\|_F^2 .$$

Objective function of L-EnsNMF

$$\left(W^{(i)},\, H^{(i)}\right) = \arg\min_{W,H\geq 0} \left\| R_L^{(i)} - WH \right\|_F^2 .$$

The main difference between *L-EnsNMF* and the (single-stage) standard NMF lies in the approach adopted to solve $W$ and $H$. That is, in standard NMF, all of $W$ and $H$ are optimized simultaneously within a single optimization framework using various algorithms such as a gradient descent

$$R^{(i)} = \begin{cases} X & \text{if} \quad i = 1 \\ \left[R^{(i-1)} - W^{(i-1)} H^{(i-1)}\right]_+ & \text{if} \quad i \geq 2 \end{cases}$$

$$R^{(i)} = \left[\left[\left[X - W^{(1)} H^{(1)}\right]_+ - W^{(2)} H^{(2)}\right]_+ \cdots - W^{(i-1)} H^{(i-1)}\right]_+ ,$$

# 3. Efficient Algorithm for Ensemble NMF

$$\left(W^{(i)}, H^{(i)}\right) = \underset{W,H\geq 0}{\arg\min} \left\| R_L^{(i)} - WH \right\|_F^2 .$$

Each sub-problem of solving $W^{(i)}$ and $H^{(i)}$ in the above equation can be represented as

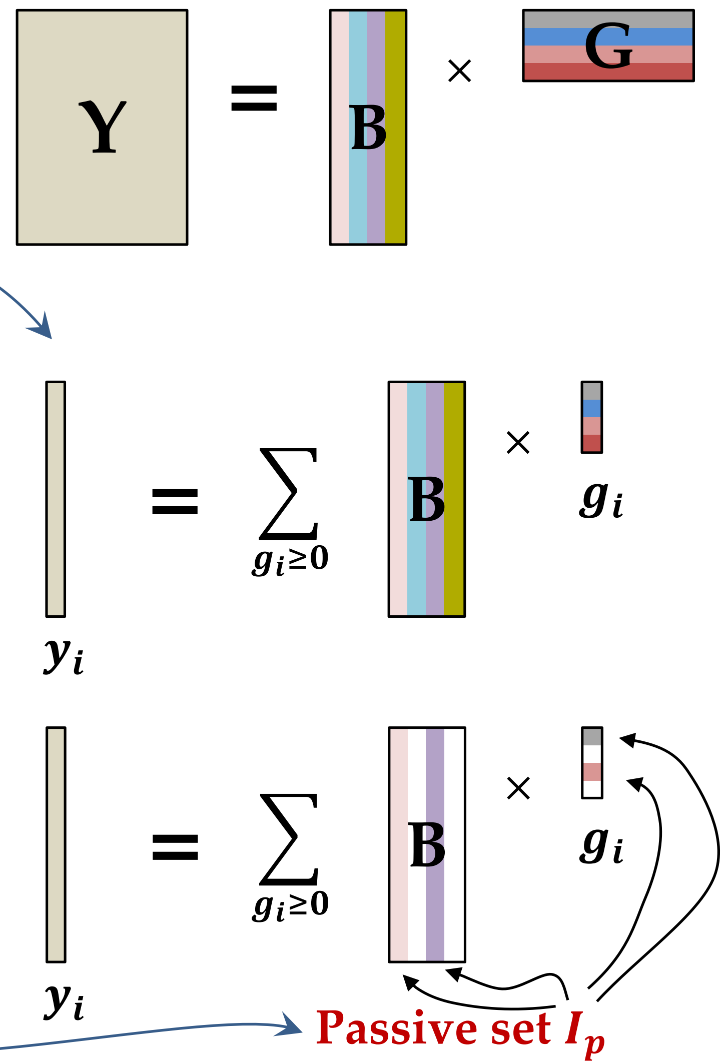
$$\min_{G\geq 0} \|Y - BG\|_F^2 = \sum_i \min_{\mathbf{g}_i \geq 0} \|\mathbf{y}_i - B\mathbf{g}_i\|_2^2 \qquad (11)$$

where $H$ is obtained by setting $B = W$, $G = H$, and $Y = X$, $W$ is obtained by setting $B = H$, $G = W$, and $Y = X^T$, and $\mathbf{g}_i$ and $\mathbf{y}_i$ are the $i$-th columns of $G$ and $Y$, respectively. Let us consider each problem in the summation operator and rewrite it as

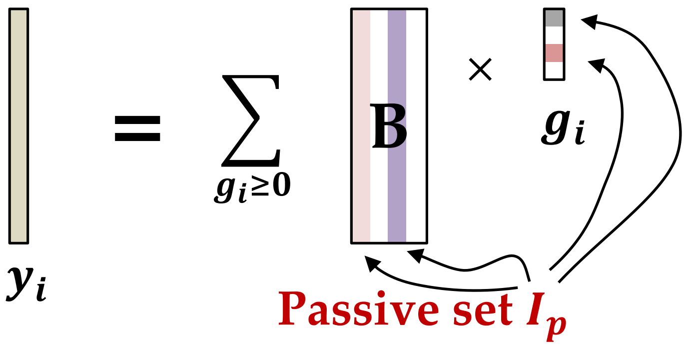
$$\min_{\mathbf{g}\geq 0} \|\mathbf{y} - B\mathbf{g}\|_2^2 , \qquad (12)$$

which is a nonnegativity-constrained least squares problem. Here, the elements of the vector $\mathbf{g}$ can be partitioned into the one containing zeros and the other containing strictly positive values, and let us call these sets of dimension indices of the active and the passive sets as $\mathcal{I}_a$ and $\mathcal{I}_p$, respectively. Once we fully know $\mathcal{I}_a$ and $\mathcal{I}_p$ for the optimal solution of Eq. (12), such an optimal solution is equivalent to the solution obtained by solving an unconstrained least squares using only the passive set of variables [20], i.e.,

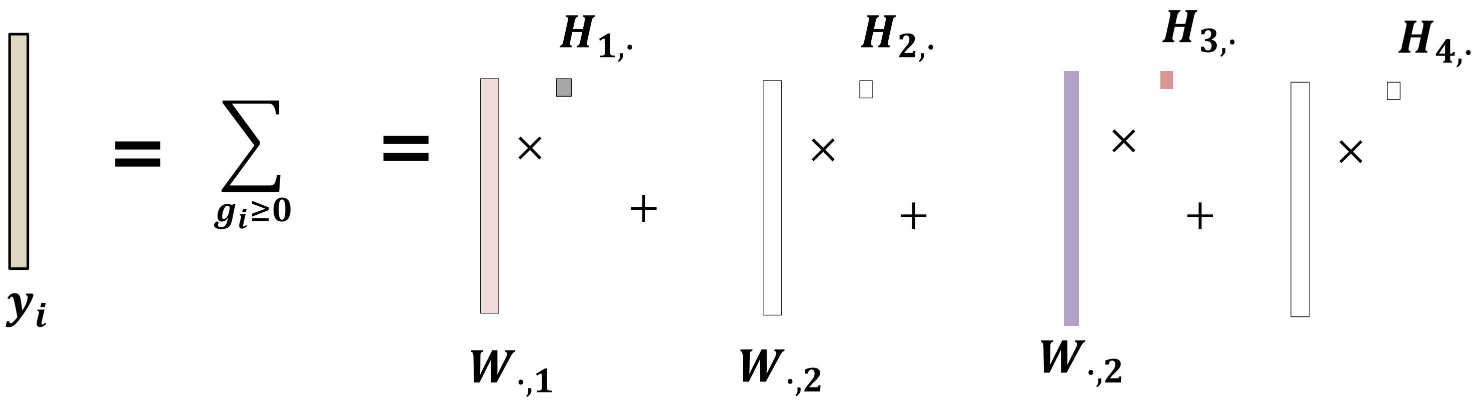
$$\min \|B(:, \mathcal{I}_p)\, \mathbf{g}_i(\mathcal{I}_p) - \mathbf{y}\|_2^2 . \qquad (13)$$

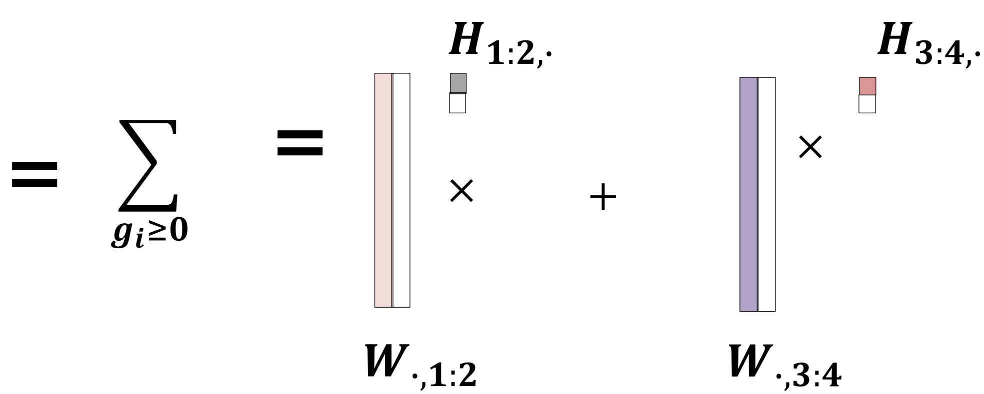


# 3. Efficient Algorithm for Ensemble NMF




The proposed model applies a greedy algorithm to exhaustively search the passive set $I_p$, which require a large time complexity




in this paper, we adopt this exhaustive search approach for an optimal active/passive set partitioning as our individual learner at each stage, which maintains the small value of $k_s$ when solving NMF at each stage.

# 4. Local Weighting

To further accelerate this process and enhance the diversity of local topics, the paper perform local weighting on the residual matrix *R(i)* so that the explained parts are suppressed while the unexplained parts are highlighte

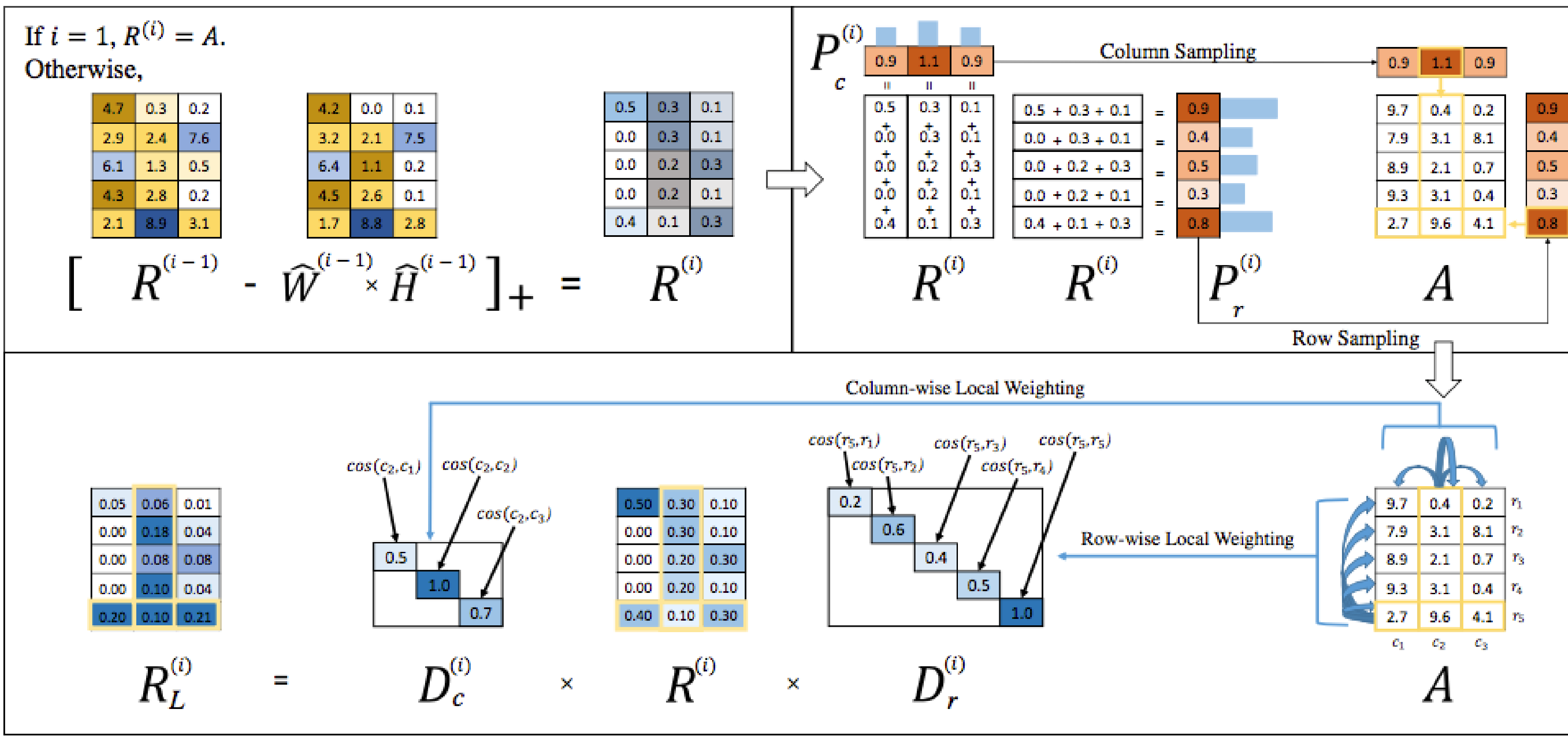
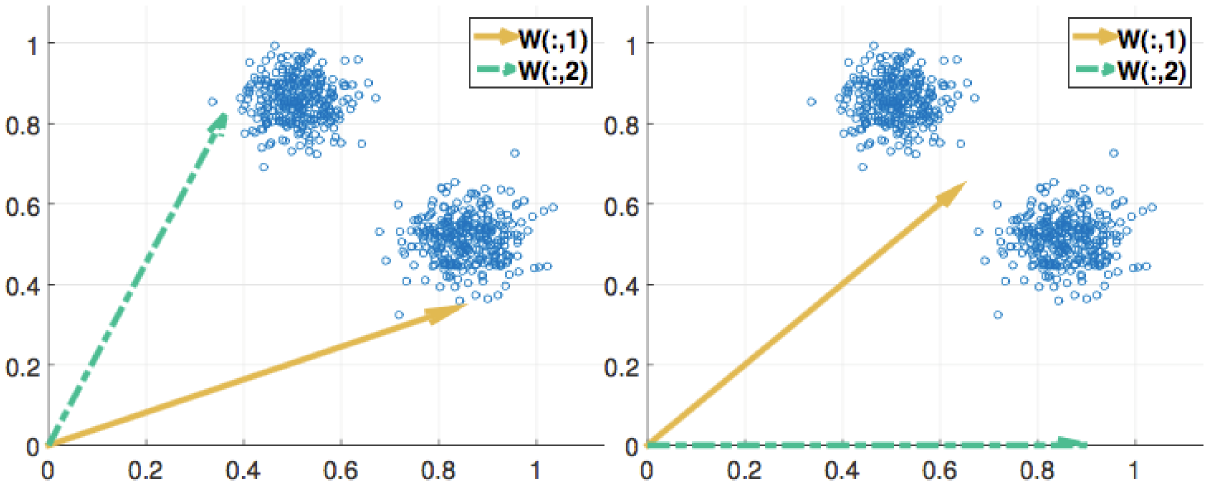



Fig. 2: Overview of the proposed ensemble approach

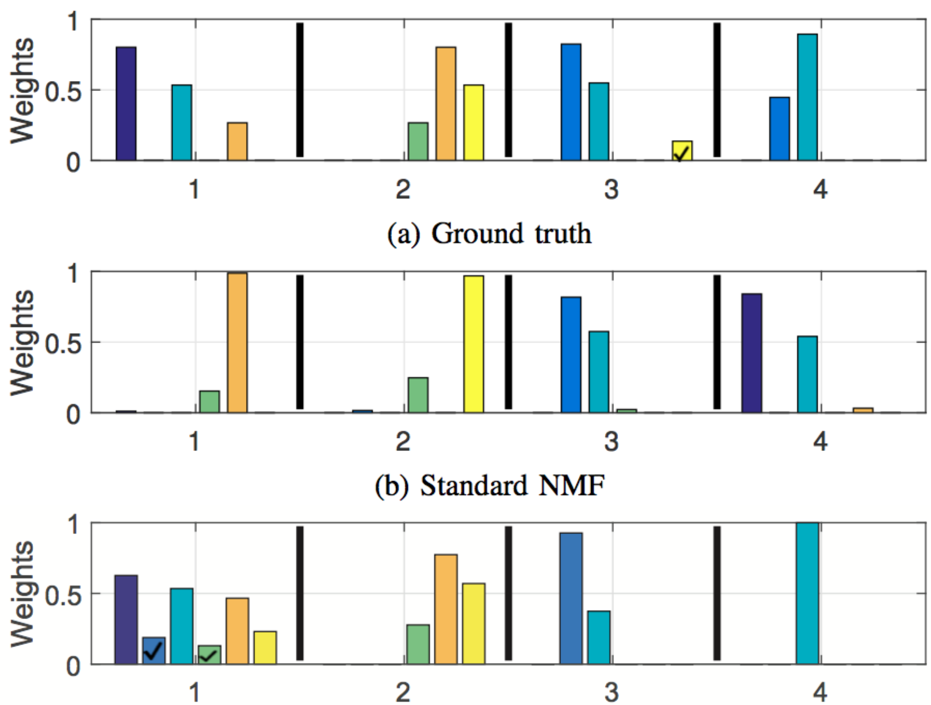
# 5. An Inexorable Issue




(a) Standard NMF (b) Deflation-based NMF

As seen in Fig. 3(a), the column vectors of $W$ generated from standard NMF in Eq. (2) successfully reveal the two components of the Gaussian mixture data. However, in the deflation approach shown in Fig. 3(b), the basis vector at the first stage, $W^{(1)} \in \mathbb{R}_+^{2\times 1}$, is computed as a global centroid and then at the second stage, $W^{(2)} \in \mathbb{R}_+^{2\times 1}$ , which is computed on the residual matrix, is shown as the vector along a single axis, $y$-axis in this case. As a result, the two bases found by the deflation-based NMF approach fail to identify the true bases. This is clearly the case where the deflation approach does not work with NMF.

# 5. An Inexorable Issue



(a) Ground truth

(b) Standard NMF

(c) Deflation-based NMF

In the case of text data, however, where the dimension is high and the matrix is highly sparse, it claims that such a deflation method can work as well as or even better than standard NMF.

The reason why the deflation-based NMF works surprisingly well with sparse high-dimensional data, e.g., text data, is because their original dimensions, e.g., keywords in text data, with large values are unlikely to overlap among different column vectors of W due to its sparsity. In this case, the deflation- based NMF can be suitable by finding these dimensions or keywords with large values in one vector at a time.

# 6. Take home massage

we introduce a novel ensemble approach of NMF for high-quality local topic discovery via a gradient boosting framework and a systematic local weighting technique. The proposed method is especially useful in disclosing local topics that are otherwise left undiscovered when using existing topic modeling algorithms.

As future work, we can add some constraint terms to the factorization framework, and also to add some prior information to steer the decomposition in a user-driven manner

*Thanks*

计算机科学与工程学院

高崇铭