Non-negative Multiple Matrix Factorization

Koh Takeuchi*, K. Ishiguro, A. Kimura, and H. Sawada,
NTT, Kyoto, Japan takeuchi.koh@lab.ntt.co.jp

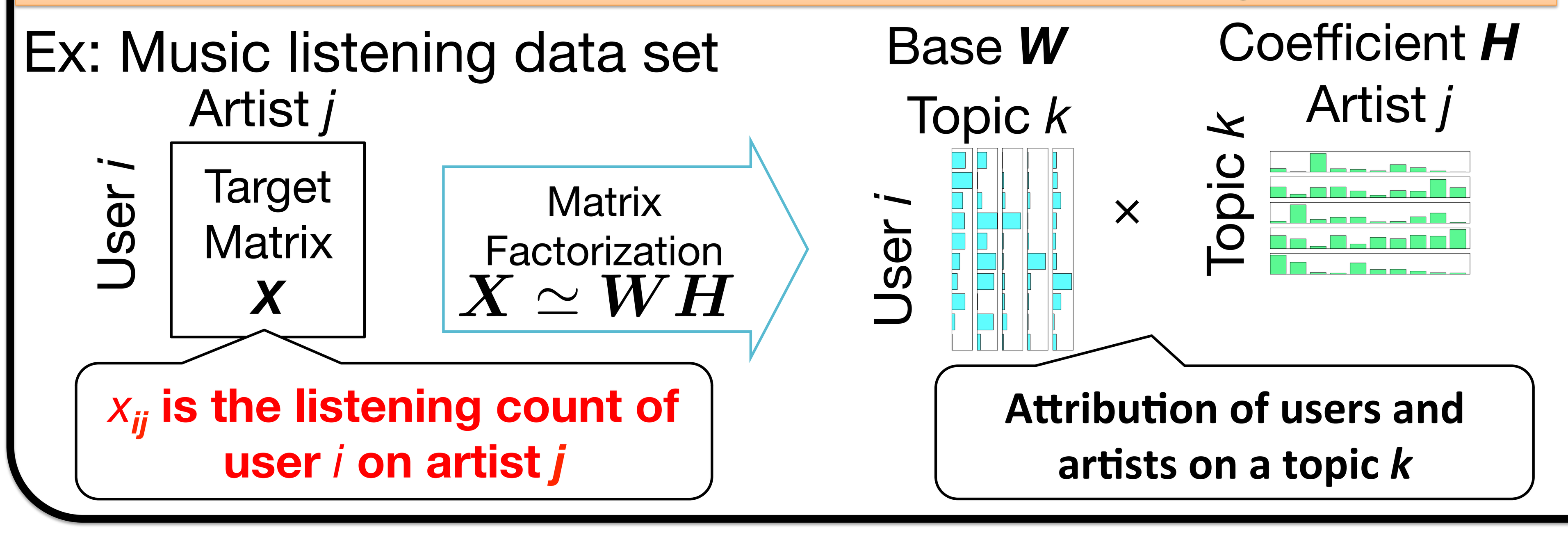
NTT IJCAI-13

Main Contribution: *We propose a novel method called Non-negative Multiple Matrix Factorization (NM2F), which integrates the information of multiple matrices and extracts common factors from the matrices*

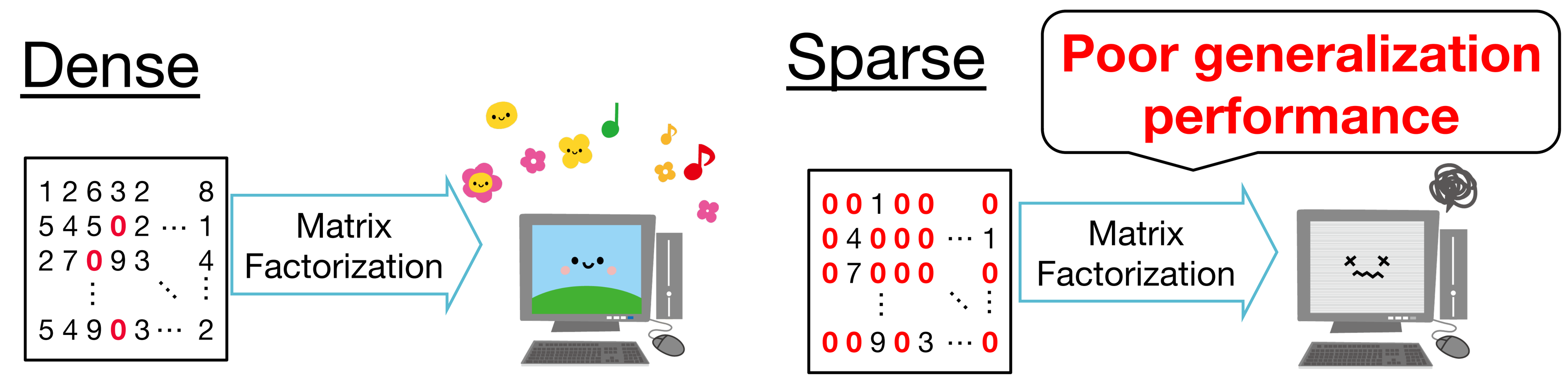
- ✓ ***Generalize Non-negative Matrix Factorization (NMF) to decompose multiple matrices***
- ✓ ***Improve generalization performance on factorizing a highly sparse target matrix***
- ✓ ***Extract common factors of the target and auxiliary matrices simultaneously***

Problem: Extract non-negative factors from a sparse matrix

NMF extracts base and coefficient factors from a target matrix ***X***



Existing methods fail when matrix X is sparse

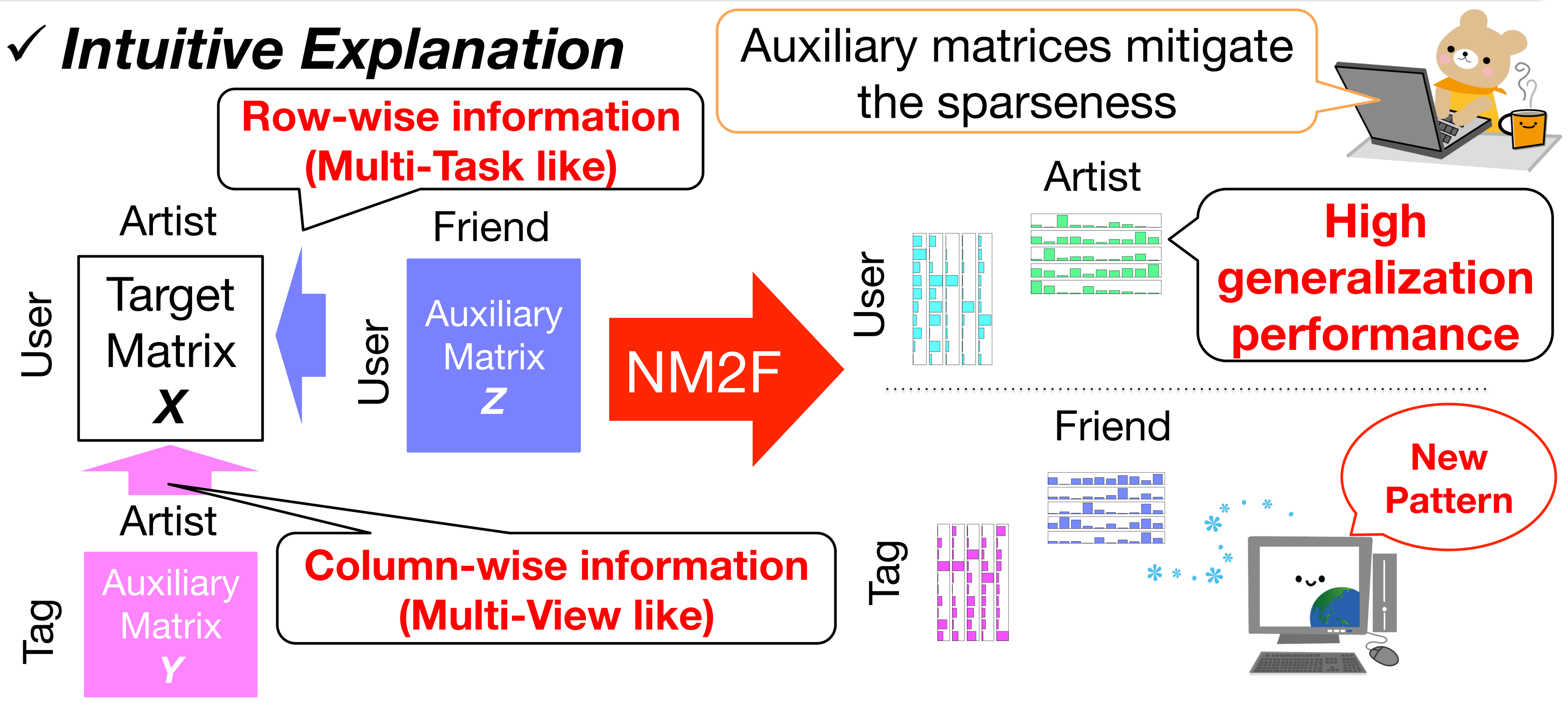


"Sparse" = Almost of elements are equal to zero

The information included in the matrix is insufficient to decompose

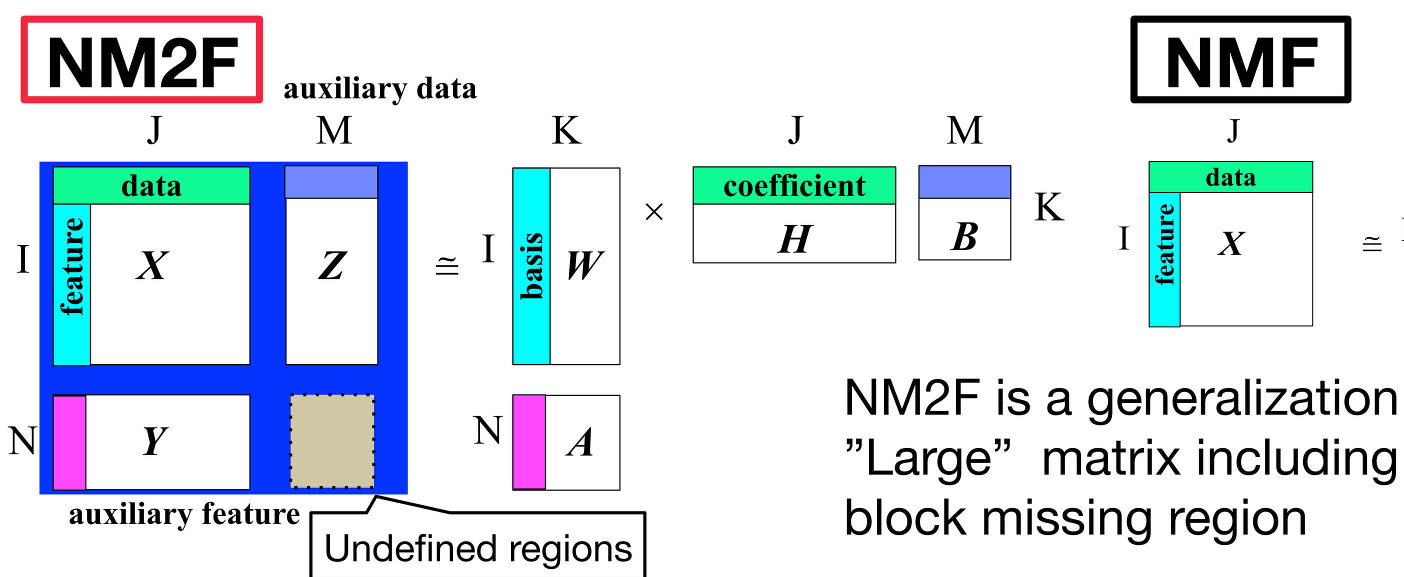
Solution: Utilize complementary data as auxiliary matrices

Auxiliary matrices share the row or column indices with the target matrix



1) Enable to factorize the highly sparse target matrix
2) Extract common factors among the target and auxiliary matrices

Mathmatical Explanation of NM2F



NM2F is a generalization of NMF, for "Large" matrix including undefined block missing region

Mathematical Details:

Objective function: minimizing reconstruction error

α and β are scaling parameters

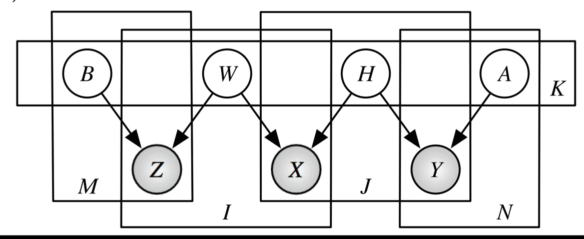
$$\min_{W,H,A,B} \mathcal{D}(X,Y,Z|W,H,A,B;\alpha,\beta) \ s.t.\ W,H,A,B \geq 0,\ \alpha,\beta \geq 0$$

$$= \mathcal{D}(X|W,H) + \alpha\mathcal{D}(Y|A,H) + \beta\mathcal{D}(Z|W,B)$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J} d(x_{i,j}|\hat{x}_{i,j}) + \alpha\sum_{n=1}^{N}\sum_{j=1}^{J} d(y_{n,j}|\hat{y}_{n,j}) + \beta\sum_{i=1}^{I}\sum_{m=1}^{M} d(z_{i,m}|\hat{z}_{i,m})$$

d is the generalized Kullback-Leibler divergence

$$\hat{x}_{i,j} = \sum_{k=1}^{K} w_{i,k}h_{k,j},\ \hat{y}_{n,j} = \sum_{k=1}^{K} a_{n,k}h_{k,j},\ \hat{z}_{i,m} = \sum_{k=1}^{K} w_{i,k}b_{k,m}$$



Experiments

Synthetic data experiment:

*Auxiliary matrices **Y** and **Z** improved average test log-likelihoods on **X** in the highly sparse situations*

	Sparse	NM2F	NMF	VBNMF	PMF
Dense	0 %	**−1.03 ± 0.09**	**−1.24 ± 0.21**	−2.72 ± 0.03	−2.47 ± 0.239
↓	9 %	**−0.99 ± 0.08**	−19.39 ± 2.82	−8.49 ± 0.60	−13.00 ± 2.74
	99 %	**−1.07 ± 0.25**	−42.45 ± 6.30	−14.55 ± 1.40	−16.25 ± 6.05
Sparse	99.9 %	**−0.86 ± 0.55**	−43.25 ± 33.45	−15.30 ± 6.30	−12.85 ± 11.20

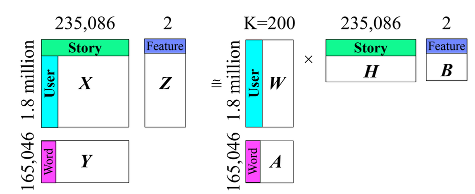
Mean and Standard deviation of test average log-likelihood by 5-fold cross validation

Real-world large data experiment:

1) Social Curation (Togetter) data set:

NM2F improved the performance (α=0.1, β=0.01)

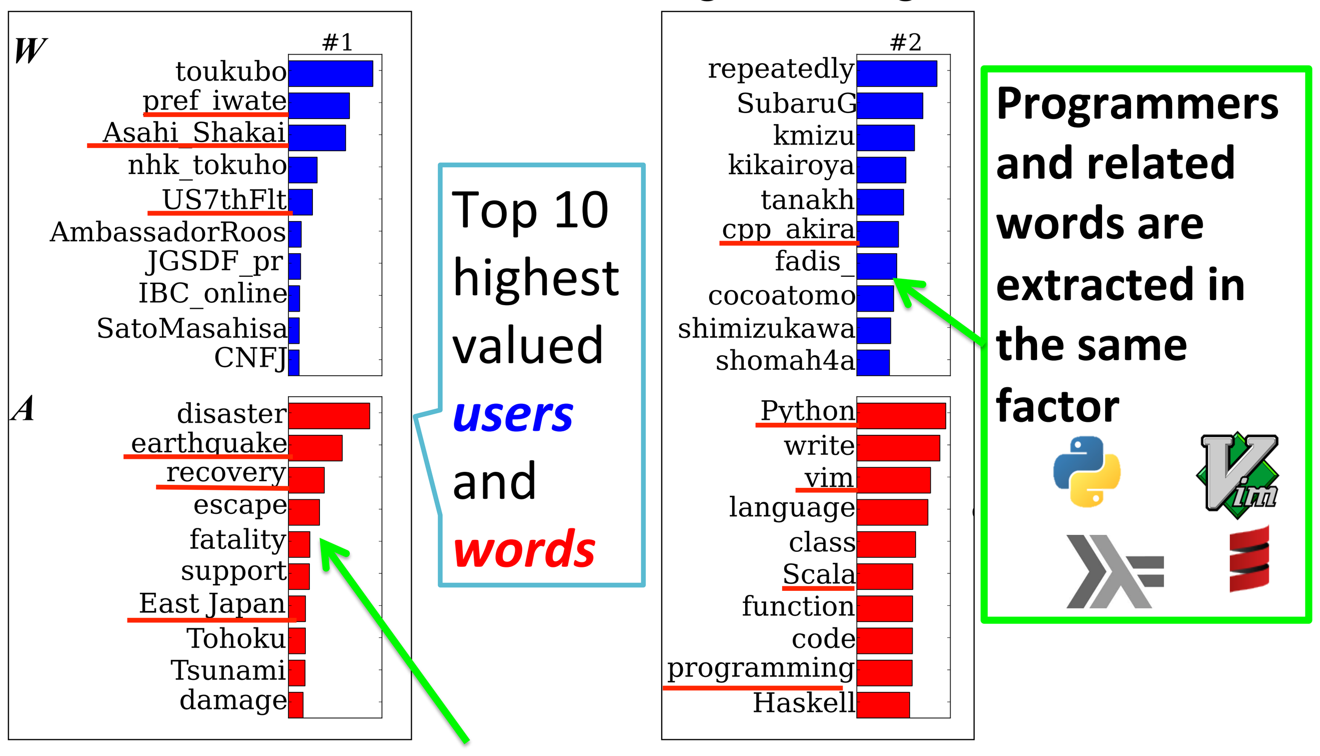
Data Set	NM2F	NMF	VBNMF	PMF
Togetter	**−12.97 ± 0.48**	−27.27 ± 0.23	N/A	N/A



1.8 Million users × 235,000 stories matrix
X is sparse (99.9982%)

NM2F Japan Tohoku Disaster Factor — Programming Factor

Ex: A story of the Japan disaster



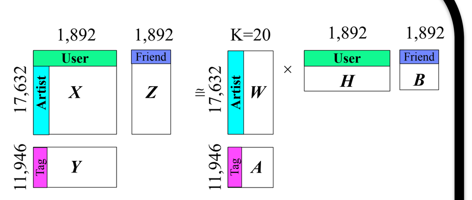
Mass media, Japan and U.S government officials who post emergency information on **the Japan disaster**

2) Last.fm data set

NM2F improved the performance (α=0.01, β=10^-5)

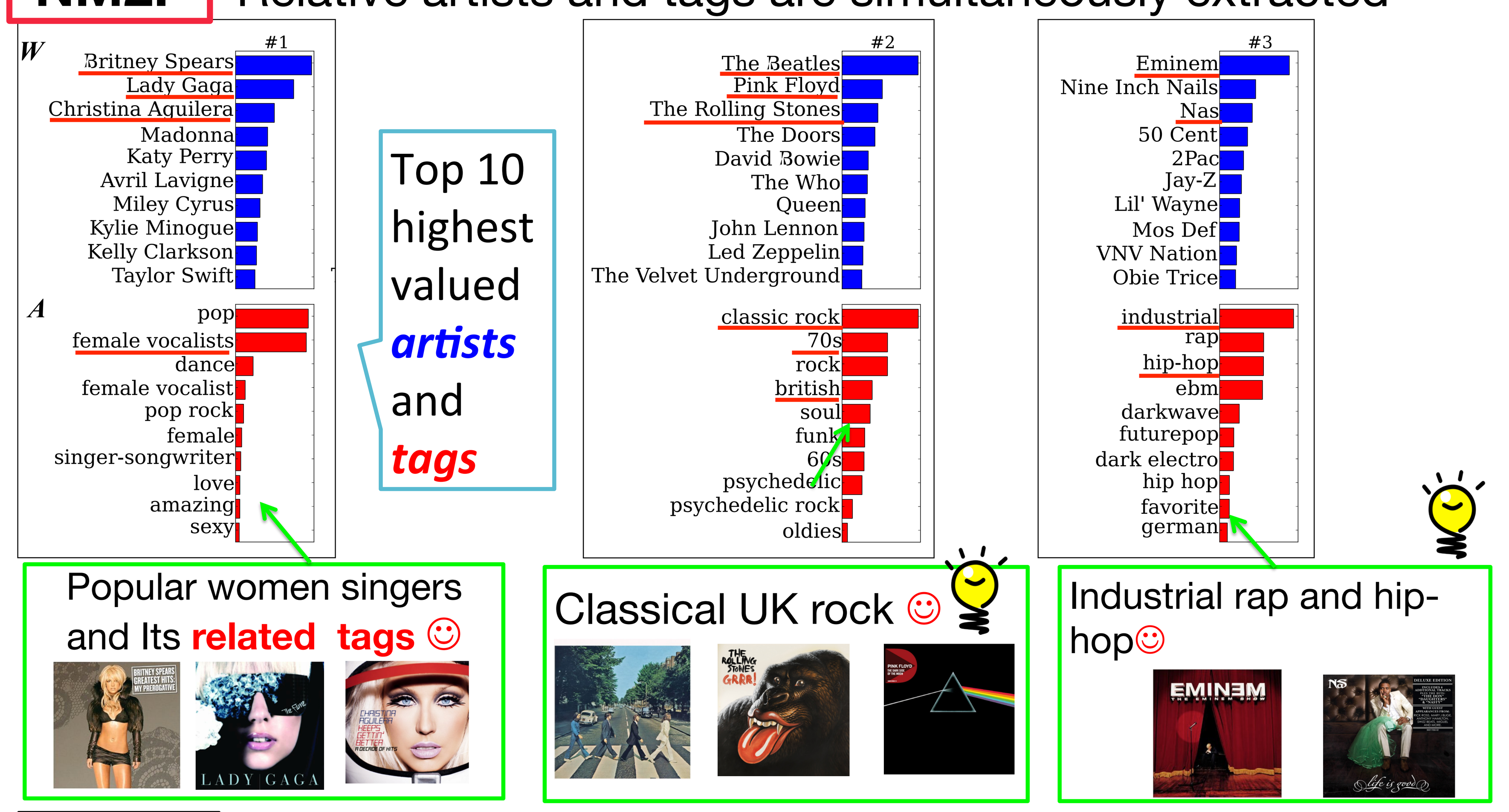
Data Set	NM2F	NMF	VBNMF	PMF
Last.fm	**−6.17 ± 0.03**	−6.90 ± 0.03	N/A	N/A

Z is less effective than the previous experiment



17,000 artists × 1,800 users matrix
X is sparse (99.75%)

NM2F Relative artists and tags are simultaneously extracted



NMF