

Non-negative Multiple Matrix Factorization

Koh Takeuchi*, K. Ishiguro, A. Kimura, and H. Sawada,

NTT, Kyoto, Japan takeuchi.koh@lab.ntt.co.jp

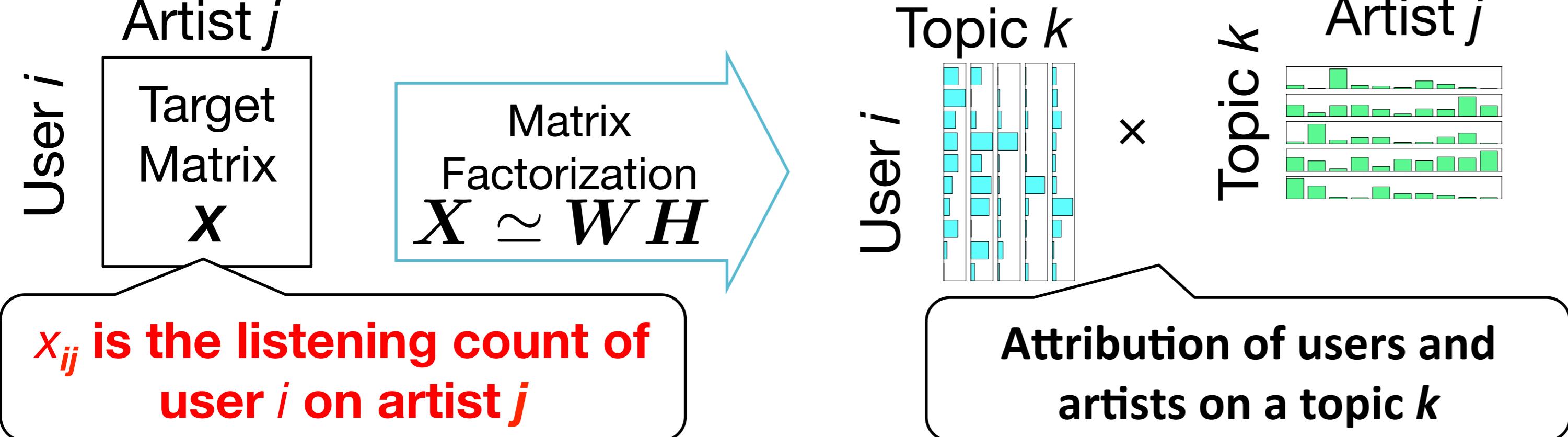
Main Contribution: We propose a novel method called Non-negative Multiple Matrix Factorization (NM2F), which integrates the information of multiple matrices and extracts common factors from the matrices

- ✓ 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

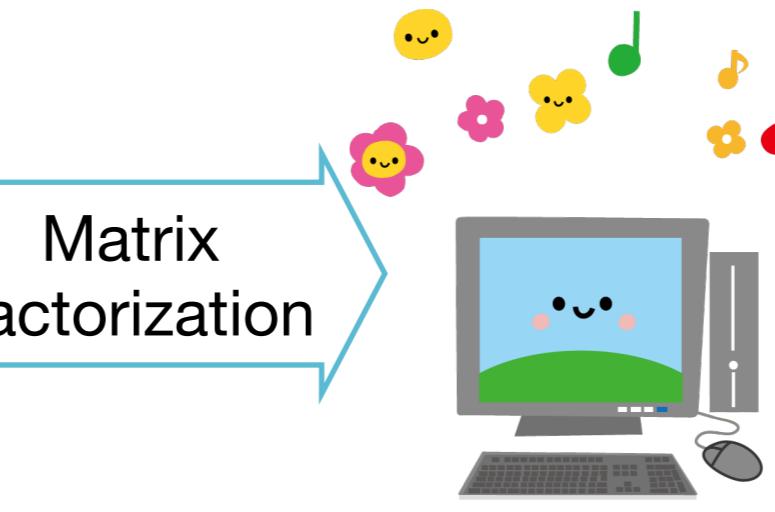
Ex: Music listening data set



Existing methods fail when matrix X is sparse

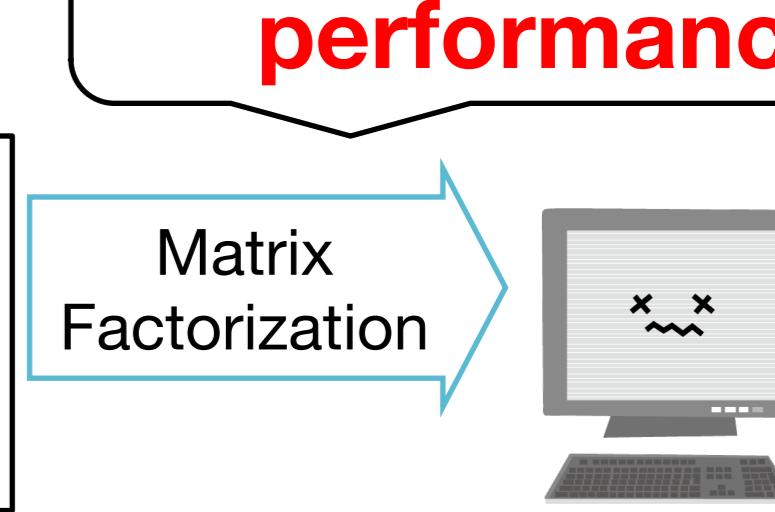
Dense

$$\begin{matrix} 1 & 2 & 6 & 3 & 2 & 8 \\ 5 & 4 & 5 & 0 & 2 & \dots & 1 \\ 2 & 7 & 0 & 9 & 3 & \dots & 4 \\ \vdots & & & & & \ddots & \\ 5 & 4 & 9 & 0 & 3 & \dots & 2 \end{matrix}$$



Sparse

$$\begin{matrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 & \dots & 1 \\ 0 & 7 & 0 & 0 & 0 & \dots & 0 \\ \vdots & & & & & \ddots & \\ 0 & 0 & 9 & 0 & 3 & \dots & 0 \end{matrix}$$



Poor generalization performance

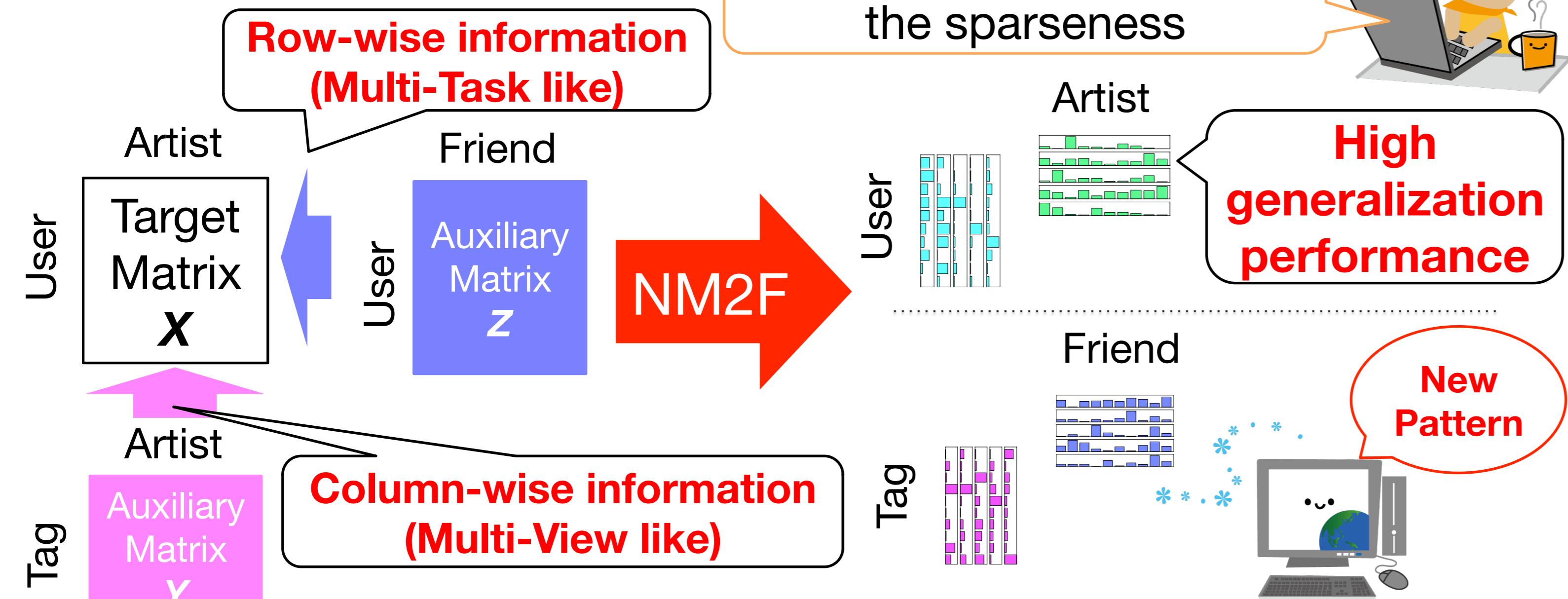
"Sparse" = Almost of elements are equal to zero

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

✓ Intuitive Explanation



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

Mathematical Explanation of NM2F

$$\text{NM2F} \quad \begin{matrix} \text{auxiliary data} \\ J \quad M \quad K \\ \text{data} \quad Z \quad W \\ I \quad \text{feature} \quad \text{basis} \\ N \quad Y \quad A \\ \text{auxiliary feature} \quad \text{undefined regions} \end{matrix} \quad \equiv \quad \begin{matrix} \text{coefficient} \\ J \quad M \quad K \\ H \quad B \\ I \quad \text{data} \quad \text{basis} \\ J \quad X \quad W \quad H \\ K \quad \text{feature} \end{matrix} \quad \equiv \quad \begin{matrix} \text{NMF} \\ K \quad J \\ W \quad H \\ J \quad \text{coefficient} \end{matrix}$$

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

Mathematical Details:

Objective function: minimizing reconstruction error

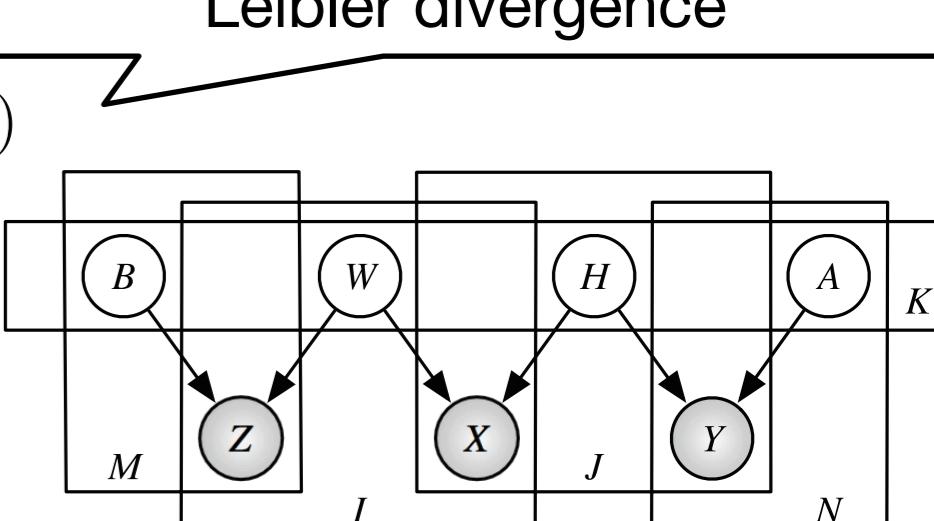
$$\min_{W,H,A,B} \mathcal{D}(X, Y, Z | W, H, A, B; \alpha, \beta) \text{ s.t. } W, H, A, B \geq 0, \alpha, \beta \geq 0$$

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

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

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

d is the generalized Kullback-Leibler divergence



Experiments

Synthetic data experiment:

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

	NM2F	NMF	VBNMF	PMF
Dense	-1.03 ± 0.09	-1.24 ± 0.21	-2.72 ± 0.03	-2.47 ± 0.239
↓	-0.99 ± 0.08	-19.39 ± 2.82	-8.49 ± 0.60	-13.00 ± 2.74
Sparse	-1.07 ± 0.25	-42.45 ± 6.30	-14.55 ± 1.40	-16.25 ± 6.05
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

$$\begin{matrix} 235.086 & 2 & K=200 \\ 165.046 & 1.8 \text{ million} & \times \\ 235.086 & 2 & \end{matrix}$$

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

2) Last.fm data set

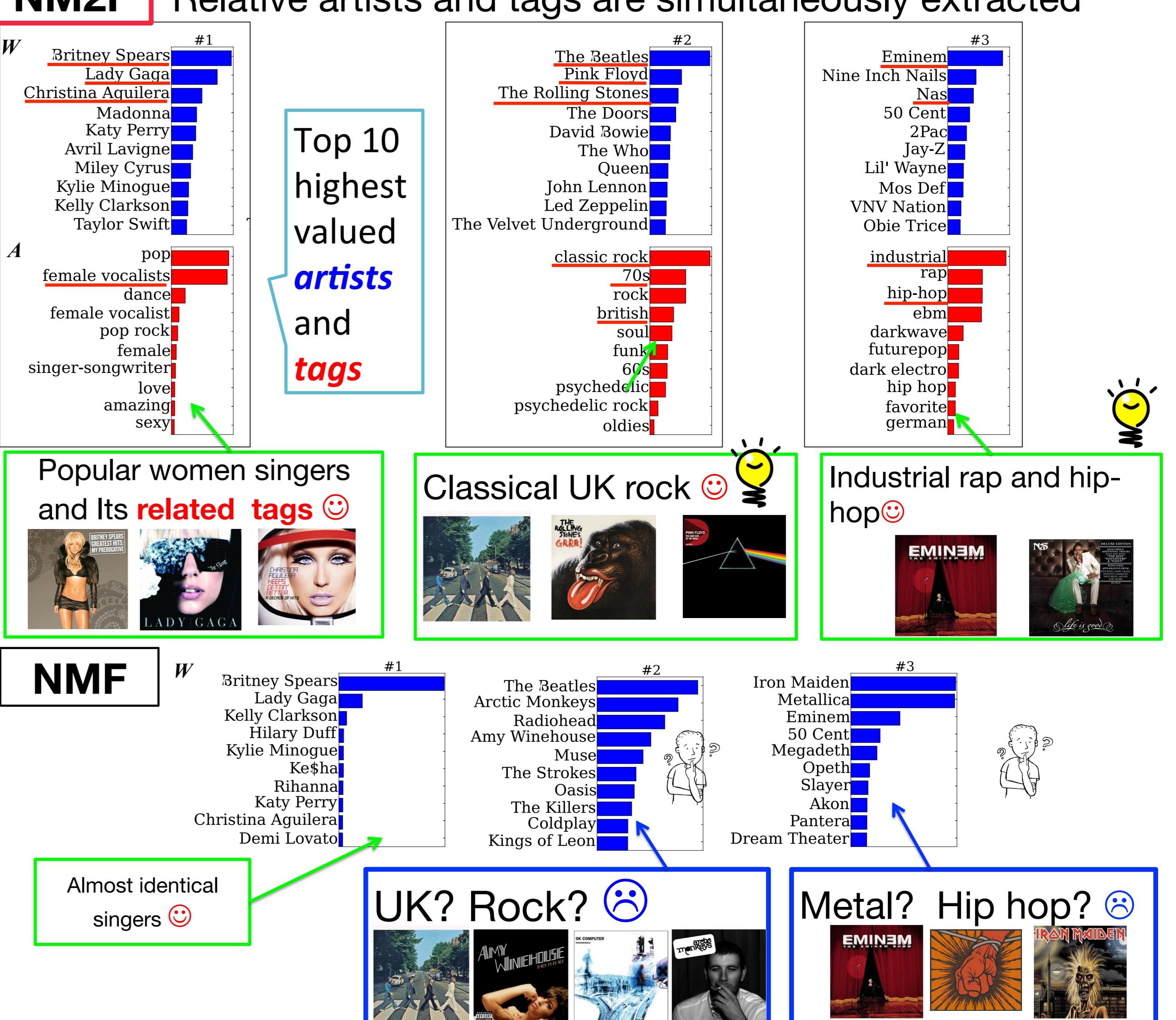
NM2F improved the performance ($\alpha=0.01, \beta=10^{-5}$)

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

NM2F

Relative artists and tags are simultaneously extracted



Real-world large data experiment:

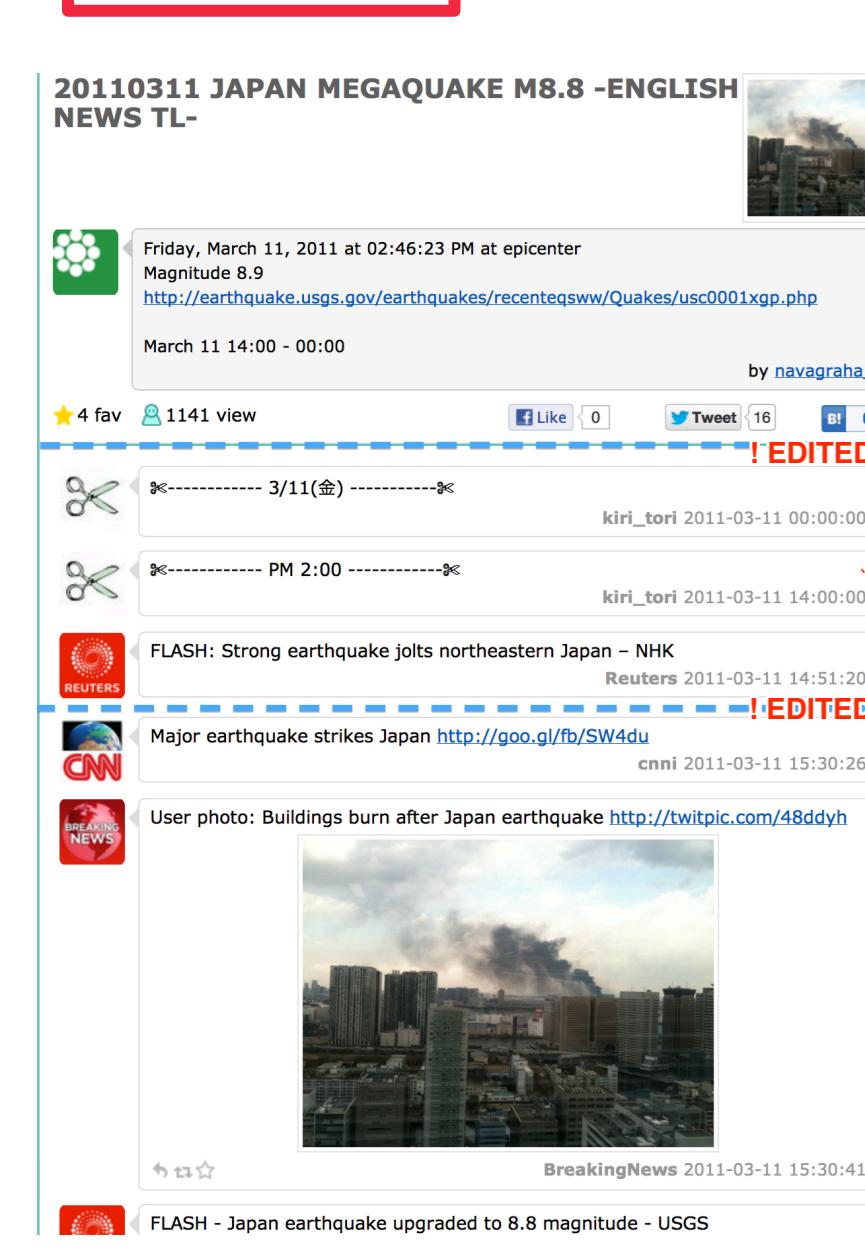
1) Social Curation (Togetter) data set:

NM2F improved the performance ($\alpha=0.1, \beta=0.01$)

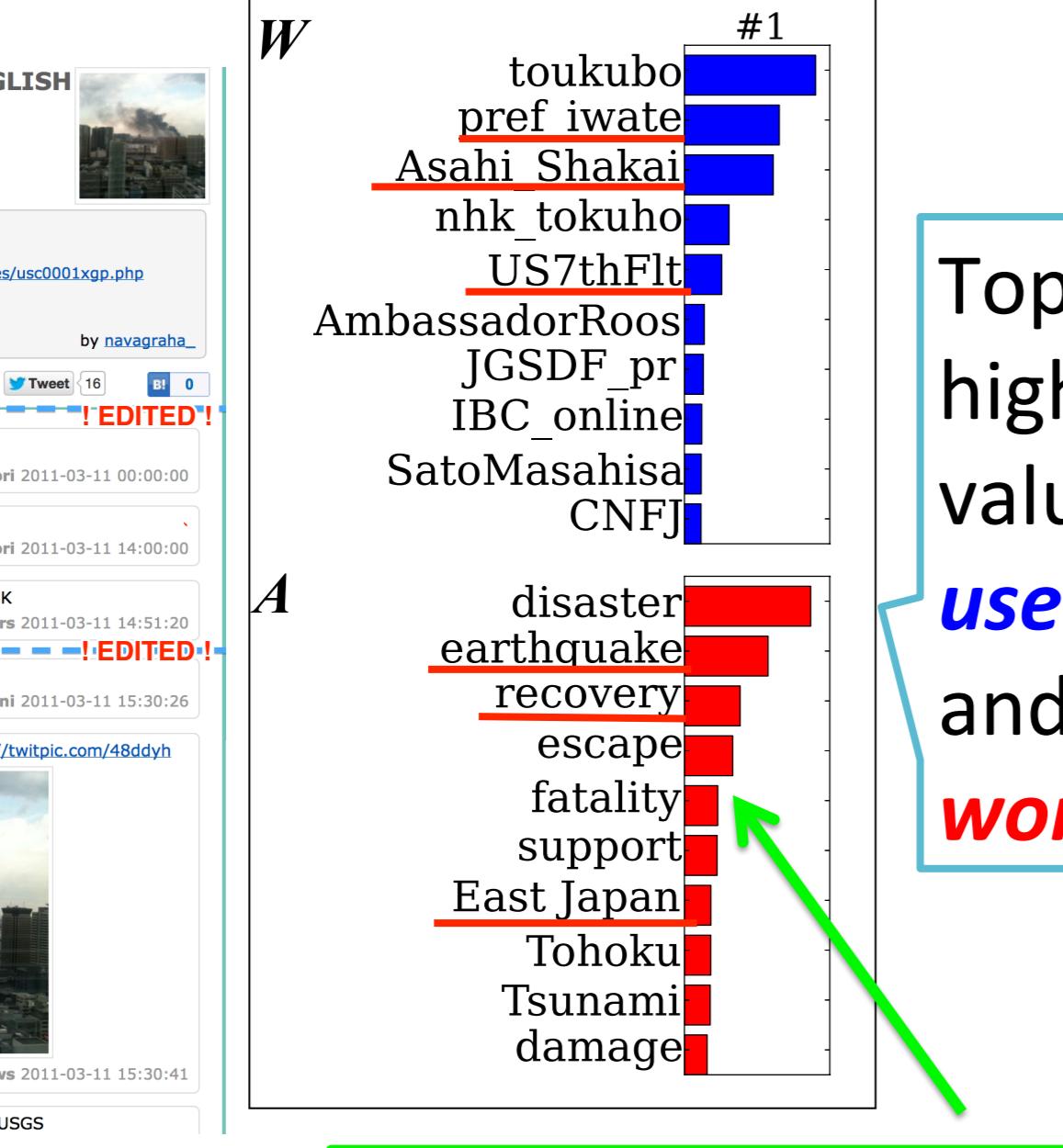
Data Set	NM2F	NMF	VBNMF	PMF
Togetter	-12.97 ± 0.48	-27.27 ± 0.23	N/A	N/A

NM2F

Japan Tohoku Disaster Factor

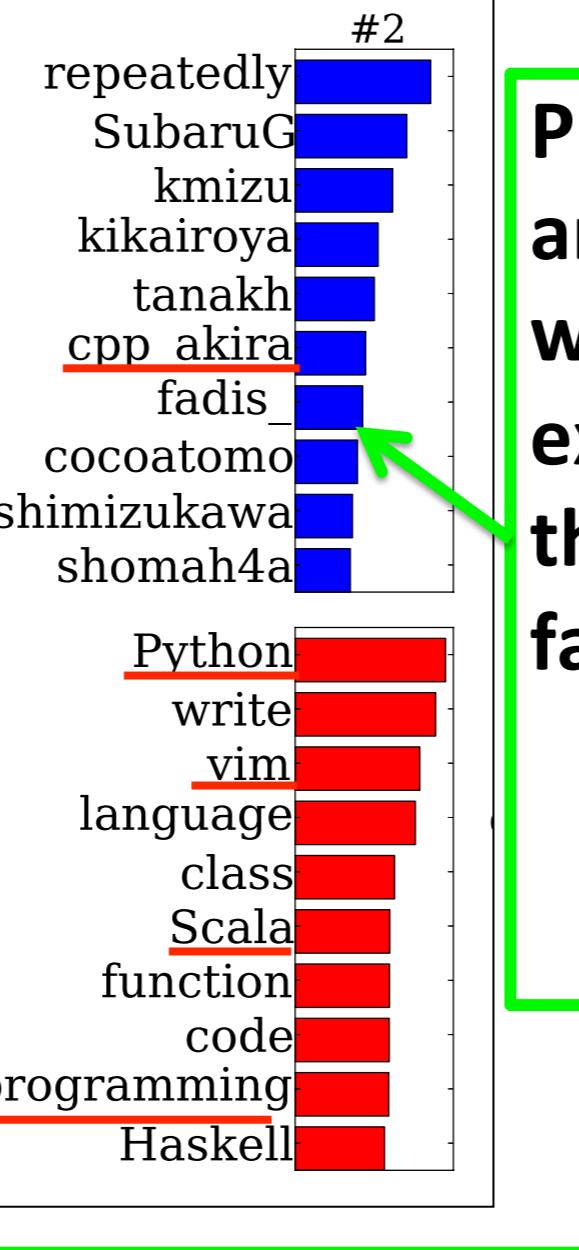


Ex: A story of the Japan disaster



Top 10 highest valued users and words

Programming Factor



Programmers and related words are extracted in the same factor

NMF

#1

Britney Spears, Lady Gaga, Christina Aguilera, Madonna, Katy Perry, Avril Lavigne, Miley Cyrus, Kylie Minogue, Kelly Clarkson, Taylor Swift

#2

The Beatles, Pink Floyd, The Rolling Stones, The Doors, David Bowie, The Who, Queen, John Lennon, Led Zeppelin, The Velvet Underground

#3

Eminem, Nine Inch Nails, Nas, 50 Cent, 2Pac, Jay-Z, Lil' Wayne, Mos Def, VNV Nation, Obie Trice

Industrial rap, hip hop, ehm, darkwave, futurepop, dark electro, hip hop, favorite german

Almost identical singers 😊

UK? Rock?

Metal? Hip hop? 😊