A Masked Mixture Model for Compact and Accurate Matrix Factorization

Yong-chan Park, Jeongyoung Lee, SeungJoo Lee, and U Kang

Data Mining Lab

Dept. of CSE

Seoul National University

Overview

- Introduction
- Proposed Method
- Experiments
- Conclusion

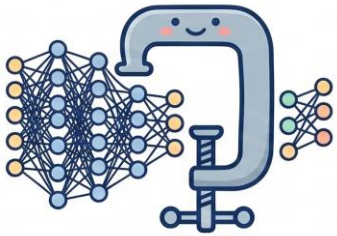
Introduction

- **Matrix Factorization (MF)**
  - MF decomposes a large matrix X into low-rank factors U and V to effectively find latent features

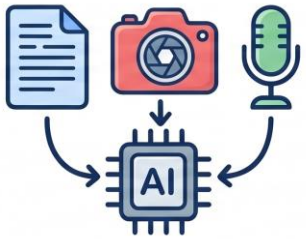
$$X \approx U V^{\top}$$

- **Applications**

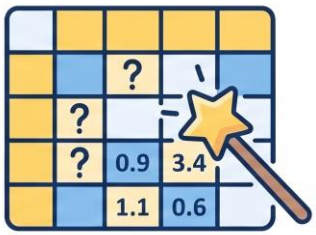
Recommender System



Model Compression



Multimodal Analysis



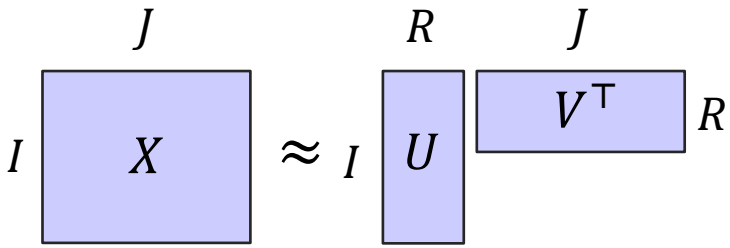
Missing Value Imputation

Introduction

- **Matrix Factorization (MF)**

$$X \approx UV^\top$$

  - $X \in \mathbb{R}^{I \times J}$, $U \in \mathbb{R}^{I \times R}$, $V \in \mathbb{R}^{J \times R}$



  - All instances use the same latent dimensions to the same extent

Key question

Should all instances use the same latent dimensions to the same extent?

Introduction

- **Masked Mixture Factorization (MMF)**
  - MMF proposes instance-wise latent masking
    - Emphasize dimensions that matter and suppress those that do not

Standard MF

All instances use
the same latent basis

MMF

Each instance gates
useful dimensions

Overview

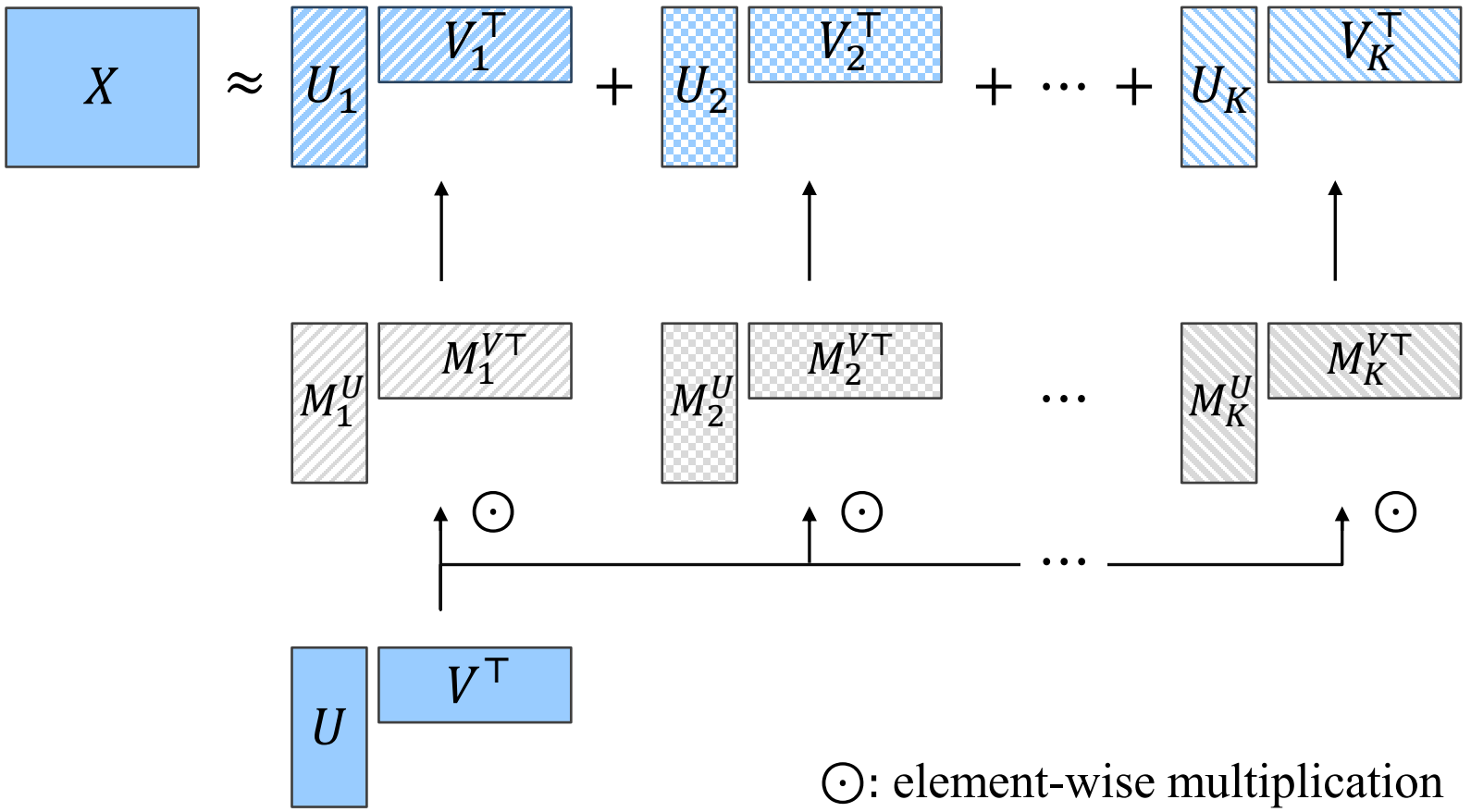
- Introduction
- **Proposed Method**
- Experiments
- Conclusion

Proposed Method

- **Challenges & Ideas**
  - **(C.1) Rigid global sharing**
    - All instances are represented in the same latent coordinate system and must rely on the same set of latent dimensions
  - **(I.1) Mixture of masked factorizations**
    - Instead of a single global factorization, we approximate the target matrix by a sum of K masked factorization components
  - **(C.2) Mask learning**
    - Learning each mask as a free matrix would significantly increase the parameter count, slow down training, and obscure interpretability
  - **(I.2) Efficient mask parameterization**
    - We restrict masks to a predefined, differentiable function family and parameterize them with a small number of learnable parameters

Proposed Method

- **Masked Mixture Factorization (MMF)**



Proposed Method

(I.1) Mixture of masked factorizations

$$X \approx \sum_{k=1}^{K} \left(U \odot M_k^U\right)\left(V \odot M_k^V\right)^\top$$

- $X \in \mathbb{R}^{I \times J}$, $U \in \mathbb{R}^{I \times R}$, $V \in \mathbb{R}^{J \times R}$, $M_k^U \in \mathbb{R}^{I \times R}$, $M_k^V \in \mathbb{R}^{J \times R}$
- K: number of masked components
- R: latent dimension (rank)
- $\odot$: element-wise multiplication

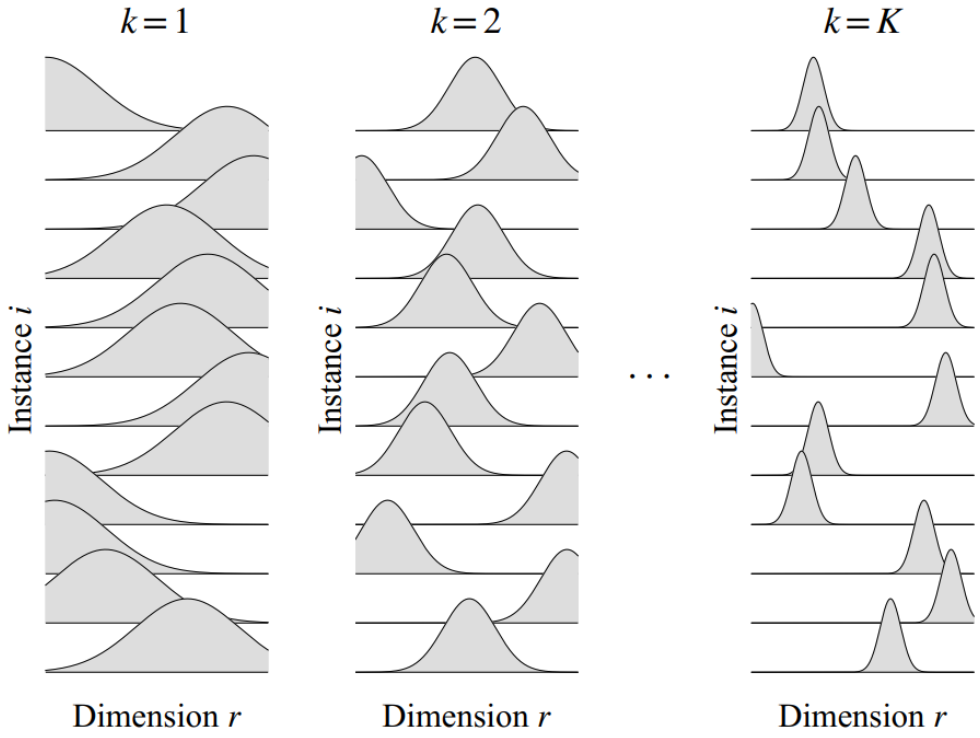
(I.2) Efficient mask parameterization

- The masks are defined by instance-wise shift parameters $\boldsymbol{s}_k^U \in \mathbb{R}^I$, $\boldsymbol{s}_k^V \in \mathbb{R}^J$ $(k = 1, \ldots, K)$, and mask functions $\{f_k^U\}_{k=1}^K, \{f_k^V\}_{k=1}^K$:

$$M_{k,i,r}^U = f_k^U\left(r; s_{k,i}^U\right) \coloneqq \frac{1}{K} f^U\left(\frac{k}{K}\left(r - s_{k,i}^U\right)\right)$$
$$M_{k,j,r}^V = f_k^V\left(r; s_{k,j}^V\right) \coloneqq \frac{1}{K} f^V\left(\frac{k}{K}\left(r - s_{k,j}^V\right)\right) \quad (1)$$

Proposed Method

- **Masked Mixture Factorization (MMF)**
  - Illustration of mask matrices (f = Gaussian)



Theoretical Analysis

- **Expressivity**
  - Rank-R MF is bounded by R
  - Under mild conditions, MMF generically reaches rank up to KR, expanding representational power under a similar budget

rank R → generic rank KR

- **Identifiability**
  - Element-wise masks break the continuous rotational symmetry of classical MF, leaving only dimension-wise scaling and producing more stable factors

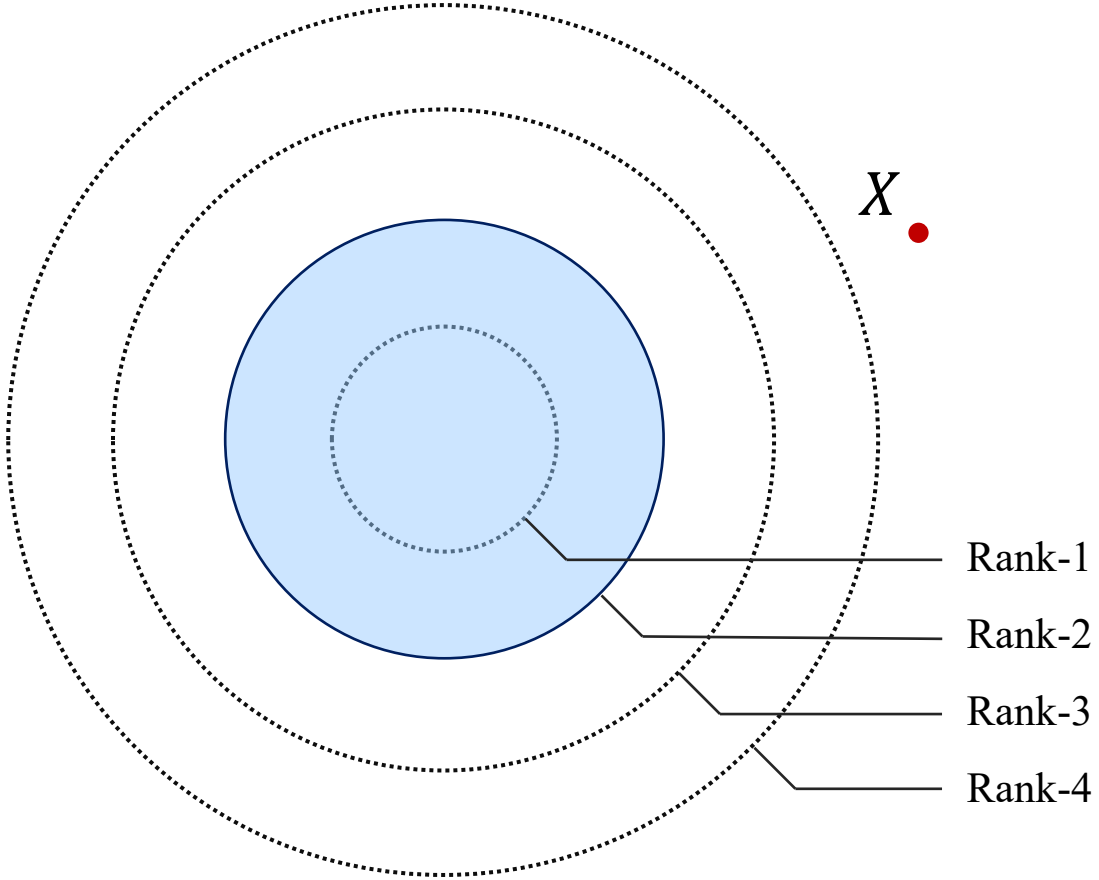
rotational symmetry ↓

Proposed Method

- **Masked Mixture Factorization (MMF)**
  - **Number of parameters**
    - **(1)** Base factors $U \in \mathbb{R}^{I \times R}, V \in \mathbb{R}^{J \times R}$ → $(I+J)R$
    - **(2)** Mask parameters $\boldsymbol{s}_k^U \in \mathbb{R}^I, \boldsymbol{s}_k^V \in \mathbb{R}^J$ → $(I+J)K$
    - **(Total)** → $(I+J)(K+R)$
  - Achieving rank KR
    - standard MF : $(I+J)KR$
    - MMF : $(I+J)(K+R)$

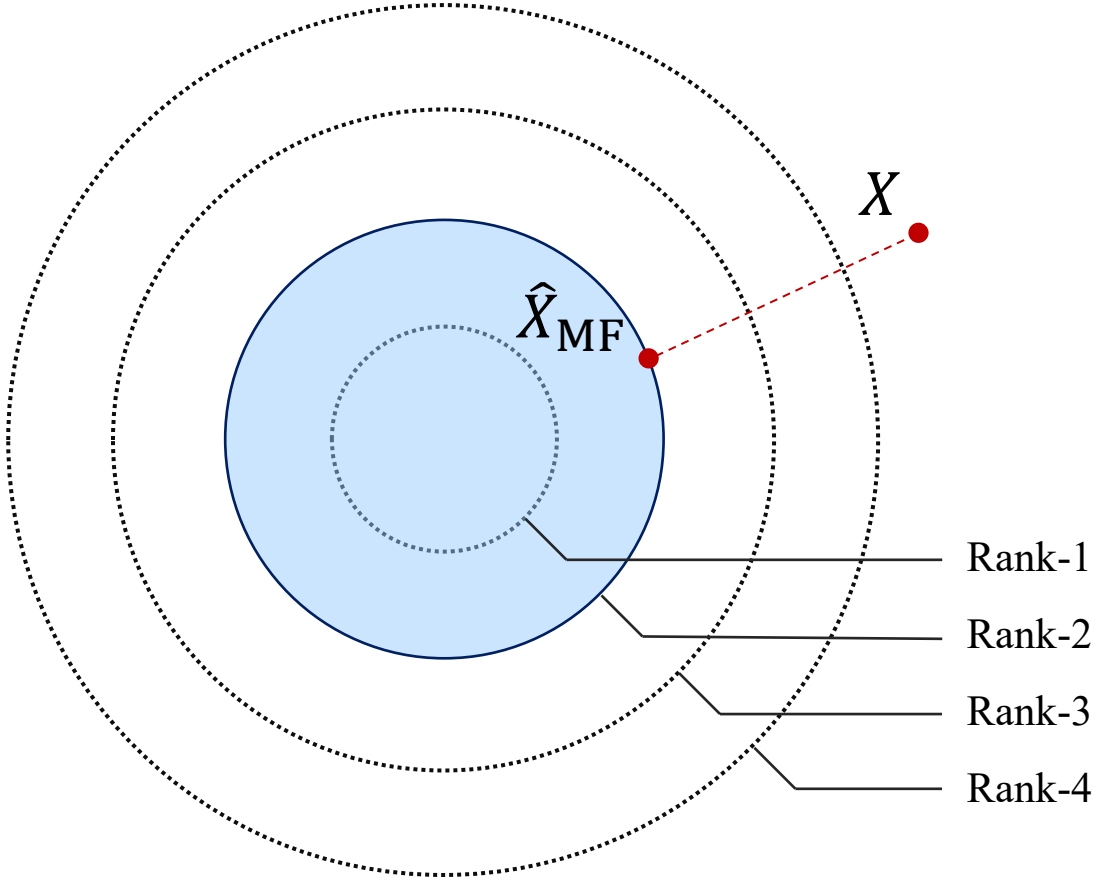
$$(I+J)(K+R) \ll (I+J)KR \quad \text{when} \quad K, R \gg 1$$

Interpretation



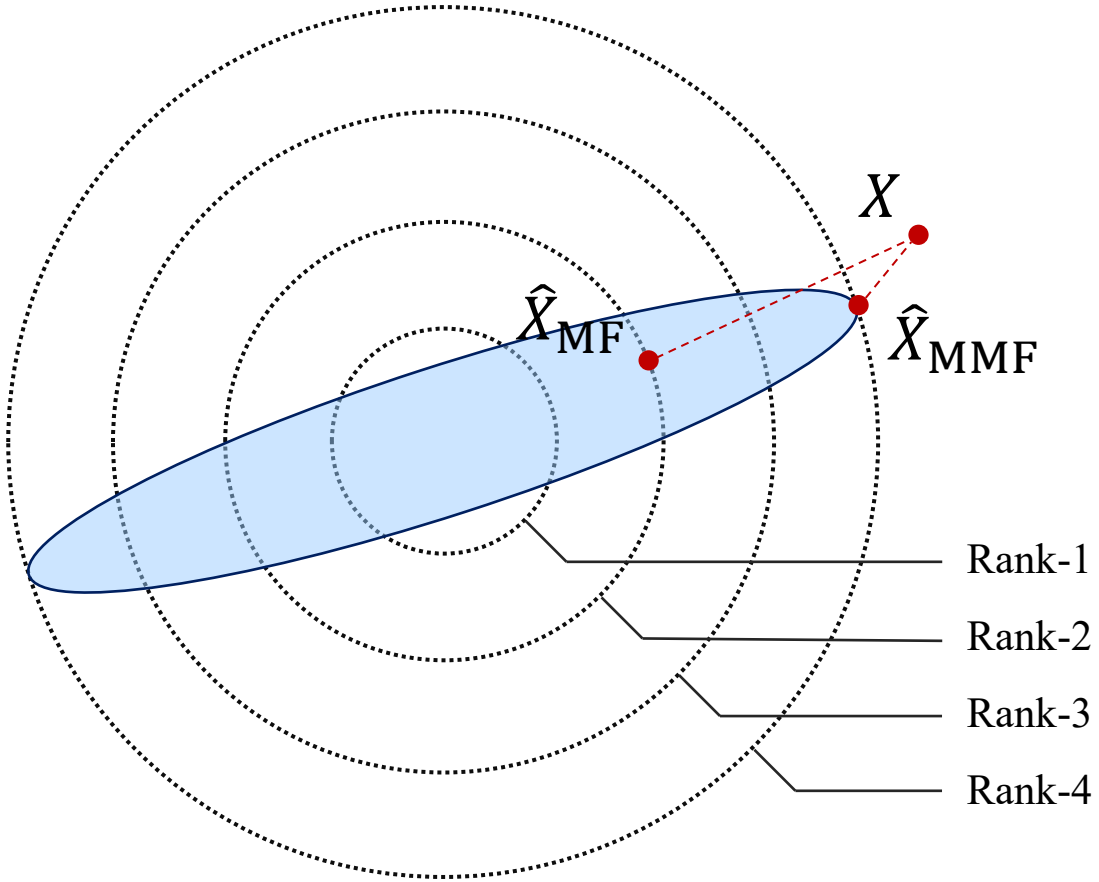
Representable set of rank = 2 models

Interpretation



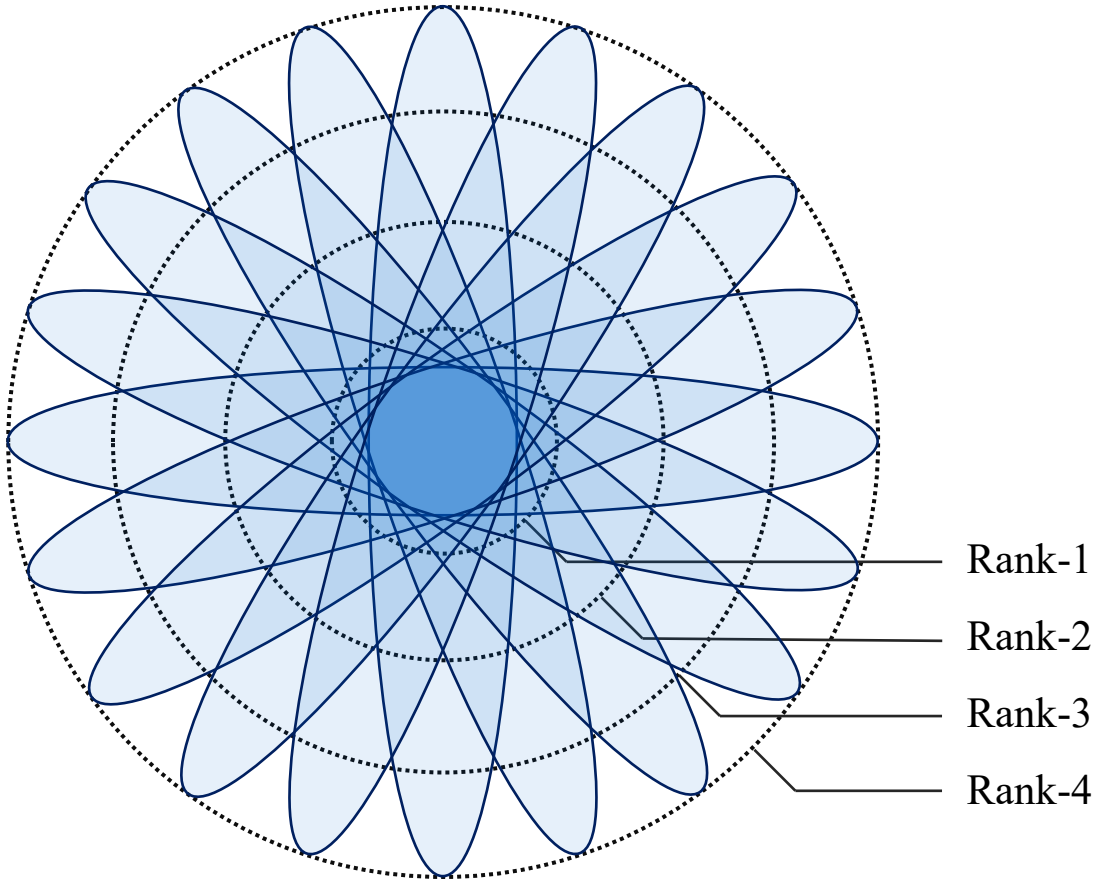
Representable set of rank = 2 models

Interpretation



Representable set of rank = 2 MMF (K=2; fixed mask)

Interpretation



Representable set of rank = 2 MMF (K=2)

Overview

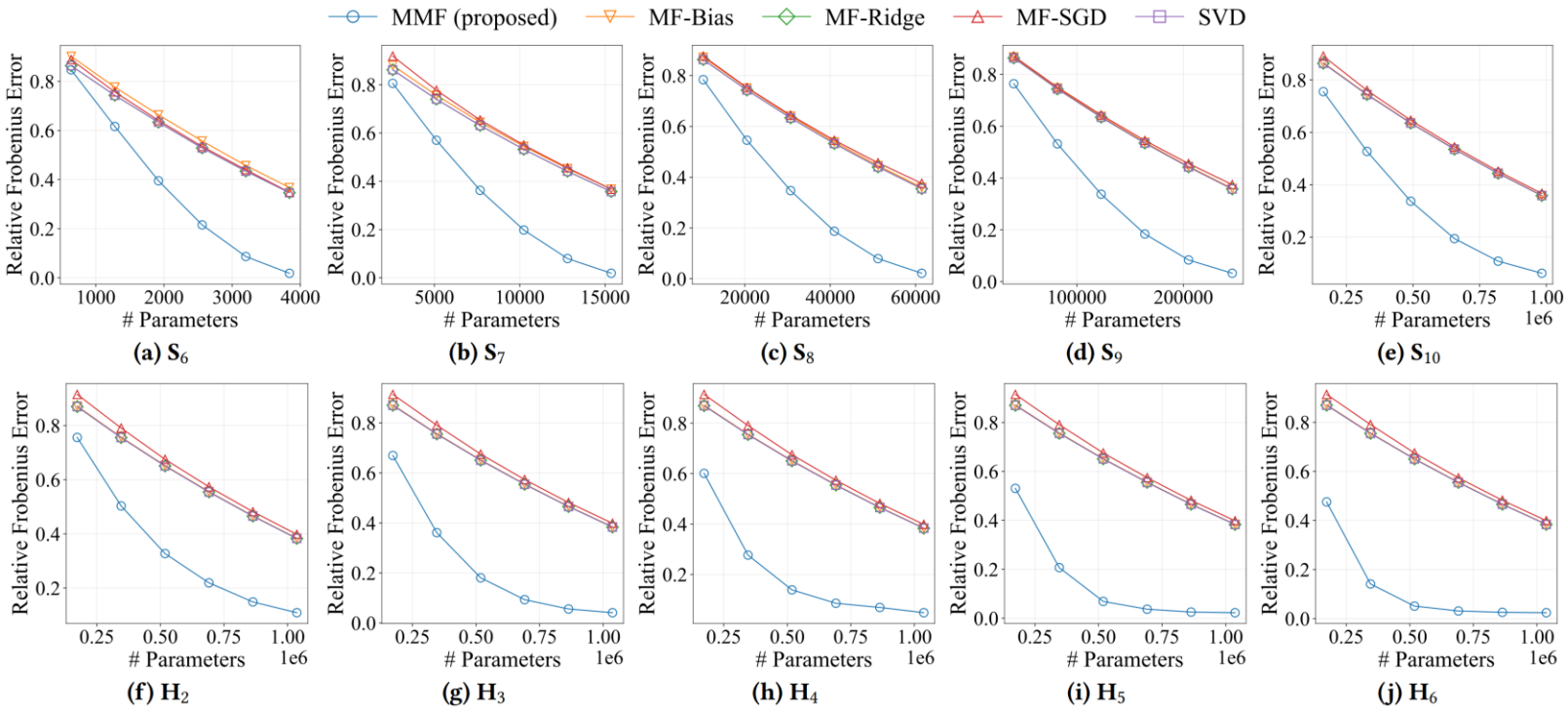
- Introduction
- Proposed Method
- **Experiments**
- Conclusion

Experiments

- **Research Questions**
  - **(Q1) Reconstruction.** Does MMF yield higher reconstruction accuracy than strong MF baselines under a matched parameter budget?
  - **(Q2) Matrix Completion.** Does MMF improve rating prediction accuracy on real-world benchmarks compared to strong baselines?
  - **(Q3) Recommendation.** Does MMF remain effective for Top-N recommendation under ranking-based evaluation?
  - **(Q4) Ablation Study.** How do mask-related hyperparameters impact the performance of MMF?
  - **(Q5) Running Time.** How fast is MMF in practice compared to MF and recent deep learning methods?
  - **(Q6) Expressivity and Identifiability.** Does MMF increase effective rank while yielding more stable and identifiable latent factors than MF?

Experiments

- **(Q1) Reconstruction**
  - Does MMF yield higher reconstruction accuracy than strong MF baselines under a matched parameter budget?



Experiments

- **(Q2) Matrix Completion**
  - Does MMF improve rating prediction accuracy on real-world benchmarks compared to strong baselines?

Method	FLIXSTER	DOUBAN	ML-100K	ML-1M	ML-10M
PMF [40]	-	0.737	0.932	0.883	-
RBM [50]	-	-	-	0.854	0.825
SVD++ [25]	-	-	0.903	0.856	-
BiasMF [26]	0.945	0.758	0.917	0.845	0.803
LLORMA [30]	-	-	-	0.833	0.782
AutoRec [52]	-	-	-	0.831	0.782
NNMF [4]	-	0.729	0.907	0.843	-
CF-NADE [67]	-	-	-	0.829	0.771
sRMGCNN [41]	0.926	0.801	0.929	0.865	0.833
⋆GC-MC [57]	0.917	0.734	0.905	0.832	0.777
⋆STAR-GCN [65]	0.879	0.727	0.895	0.832	0.770
IGMC [66]	**0.872**	**0.721**	0.905	0.857	O.O.M.
⋆IDCF-GC [61]	0.910	0.733	0.893	0.835	O.O.M.
IMC-GAE [53]	0.884	**0.721**	0.897	0.829	O.O.M.
⋆GHRS [3]	-	-	0.887	0.833	-
⋆MoRGH [68]	-	-	**0.881**	**0.827**	-
MMF (proposed)	0.898	0.726	0.884	**0.827**	**0.769**

⋆ denotes methods using side information.

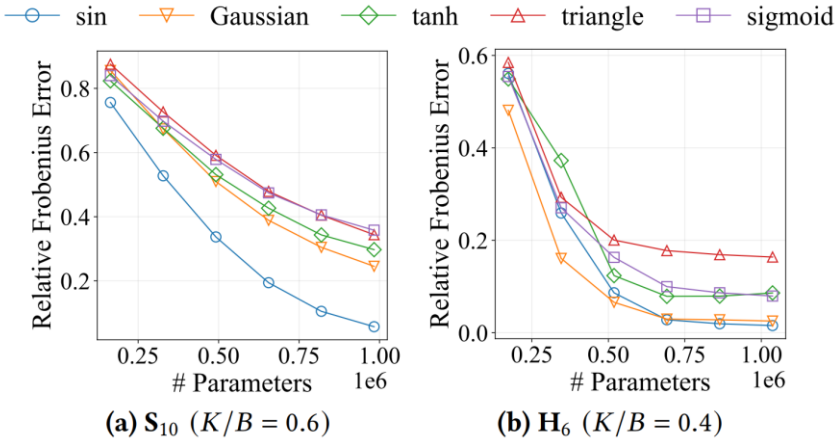
Experiments

- **(Q3) Recommendation**
  - Does MMF remain effective for Top-N recommendation under ranking-based evaluation?

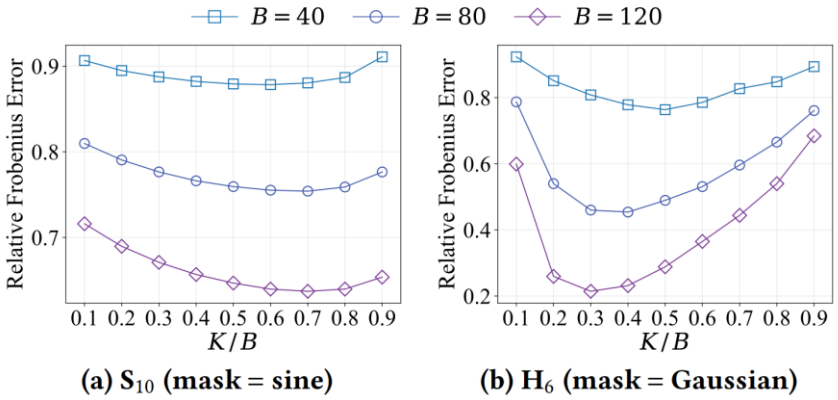
Method	Recall@10	NDCG@10	Recall@20	NDCG@20
BPRMF [48]	0.1745	0.2410	0.2624	0.2516
NeuMF [10]	0.1525	0.2183	0.2352	0.2285
NGCF [58]	0.1761	0.2467	0.2677	0.2576
LightGCN [9]	0.1782	0.2501	0.2699	0.2606
SGL [59]	0.1804	**0.2535**	0.2732	0.2644
MMF (proposed)	**0.1807**	0.2530	**0.2742**	**0.2647**

Experiments

- **(Q4) Ablation Study**
  - How do mask-related hyperparameters impact the performance of MMF?



Effect of mask families on MMF's performance



Impact of mask count on performance under a fixed parameter budget

Experiments

- **(Q5) Running Time**
  - How fast is MMF in practice compared to MF and recent deep learning methods?

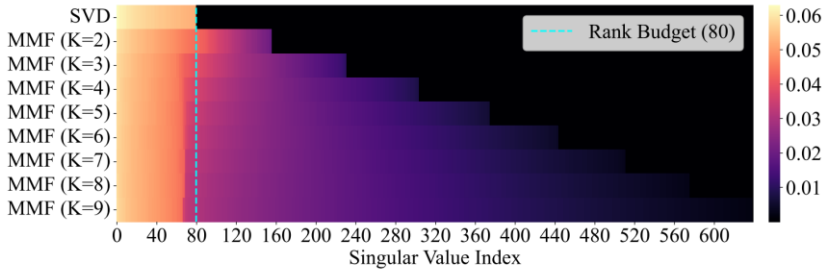
Method	S_6	S_7	S_8	S_9	S_{10}
MMF ($K/B = 0.1$)	1.118	1.129	1.163	1.195	1.264
MMF ($K/B = 0.2$)	1.062	1.076	1.091	1.120	1.213
MMF ($K/B = 0.3$)	1.174	1.200	1.221	1.288	1.295
MMF ($K/B = 0.4$)	1.209	1.211	1.228	1.293	1.344
MMF ($K/B = 0.5$)	1.223	1.218	1.231	1.206	1.301
MMF ($K/B = 0.6$)	1.215	1.222	1.233	1.204	1.299
MMF ($K/B = 0.7$)	1.215	1.229	1.237	1.285	1.295
MMF ($K/B = 0.8$)	1.226	1.213	1.242	1.280	1.292
MMF ($K/B = 0.9$)	1.242	1.232	1.240	1.270	1.274
MF-Bias	0.902	0.914	0.915	0.928	0.967
MF-Ridge	0.857	0.886	0.889	0.897	0.929
MF-SGD	0.529	0.592	0.594	0.605	0.682
SVD*	4.504	9.008	23.18	43.03	84.07

*For SVD, the total running time is reported.

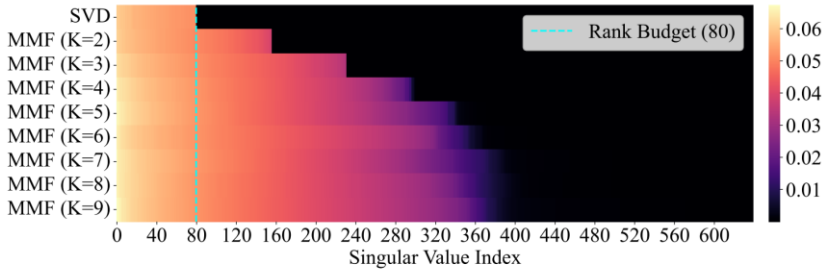
	ML-100K		ML-1M	
Method	Training	Inference	Training	Inference
sRMGCNN [41]	46.969	0.0116	407.80	0.0221
GC-MC [57]	43.597	0.0071	384.26	**0.0159**
IGMC [66]	984.03	13.682	7467.8	54.861
IDCF-GC [61]	74.389	0.1016	876.45	4.0528
IMC-GAE [53]	58.942	0.0153	412.11	0.0312
MMF	**15.158**	**0.0052**	**67.340**	0.0448

Experiments

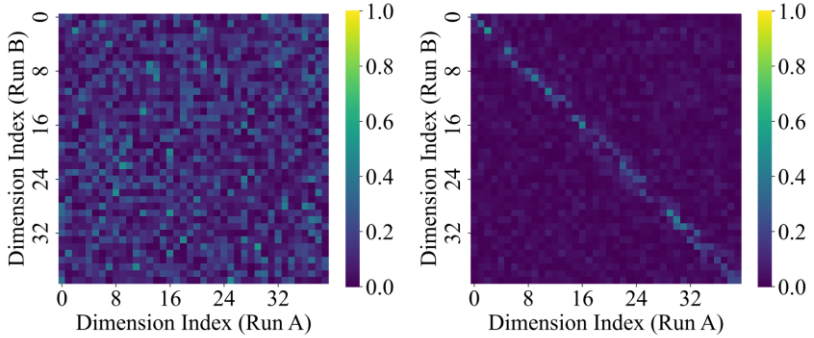
- **(Q6) Expressivity and Identifiability**
  - Does MMF increase effective rank while yielding more stable and identifiable latent factors than MF?



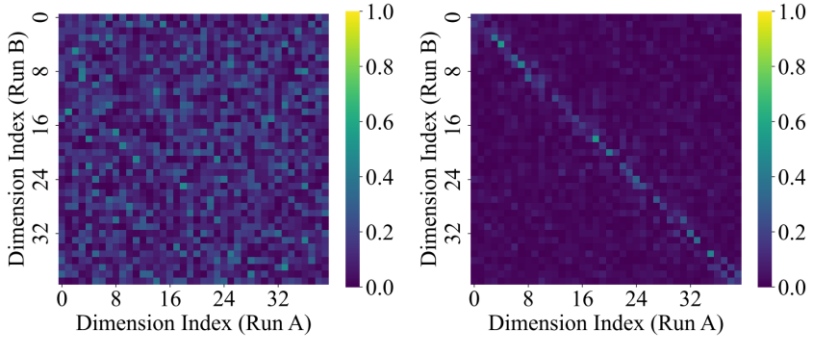
(a) S_{10} (Random Dense Matrix)



(b) H_6 (Heterogeneous Block-Diagonal Matrix)



(a) Standard MF on S_{10} (b) MMF on S_{10}



(c) Standard MF on H_6 (d) MMF on H_6

Overview

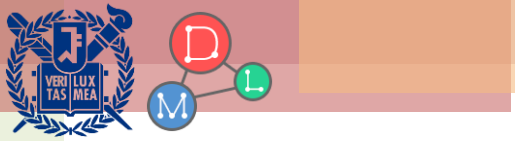
- Introduction
- Proposed Method
- Experiments
- **Conclusion**

Conclusion

- **Masked Mixture Factorization (MMF)**
  - MMF proposes instance-wise latent masking
    - Emphasize dimensions that matter and suppress those that do not

Standard MF	MMF
All instances use the same latent basis	Each instance gates useful dimensions

- **Contributions**
  - ✓ Methodology: masked mixture of shared factors
  - ✓ Theory: expressivity + identifiability guarantees
  - ✓ Experiments: reconstruction, completion, ranking



THANK YOU!

https://github.com/snudatalab/MMF

ycpark6767@gmail.com