contact: evgeny.frolov@skoltech.ru
read more: https://www.eigentheories.com

HybridSVD

When Collaborative Information is Not Enough

Evgeny Frolov[1], Research Scientist
Ivan Oseledets[1,2], Associate Professor

[1]Skolkovo Institute of Science and Technology
[2]Institute of Numerical Mathematics of the Russian Academy of Sciences

1. BACKGROUND

Proper tuning of simpler baselines vs. modern SOTA shows surprising results.
Part of the reason – tuning can be hard and cumbersome!
Can we reduce tuning hassles at least for the baselines?

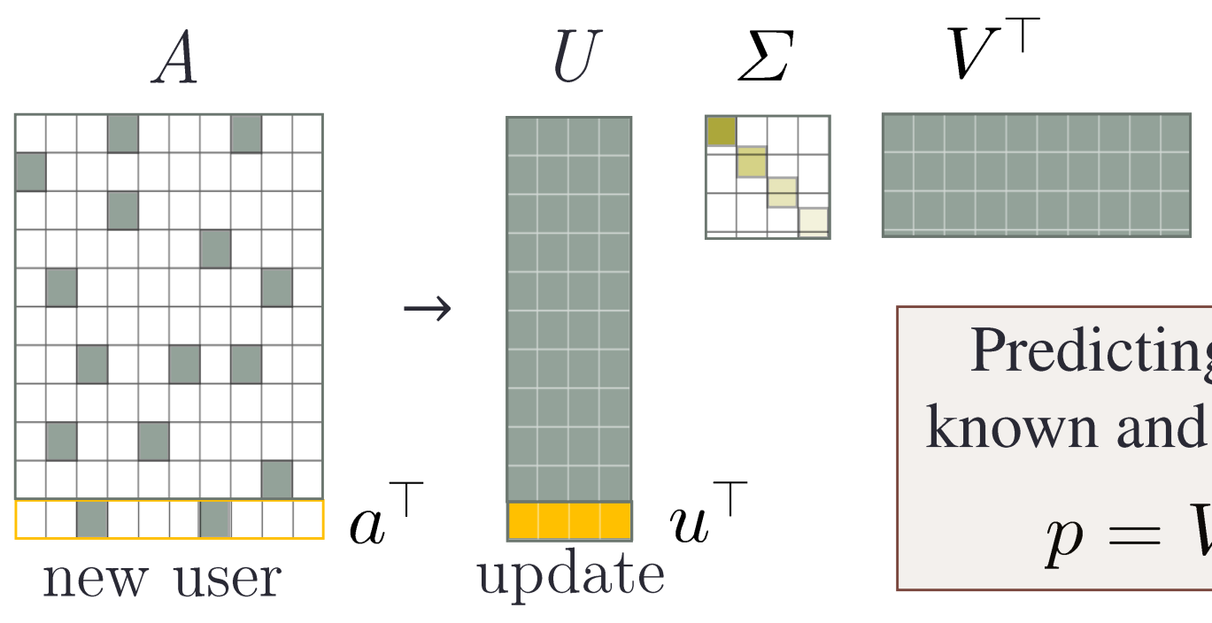
Example: PureSVD $\|A - U\Sigma V^\top\|_F^2 \to \min$

Advantages:

- ✓ simple tuning via rank truncation
- ✓ minimal storage requirements
- ✓ supports real-time & session-based recs
- ✓ stable, deterministic output
- ✓ highly optimized implementations
- ✓ scales to ~**billion-size** problems

e.g., https://github.com/criteo/Spark-RSVD

Interesting fact: the 3rd most cited work (via Scopus) among all ACM RecSys conference proceedings papers!



Predicting for both known and new users:
$p = VV^\top a$

Data-debiasing trick [Nikolakopoulos et al. 2019]:

$$A \leftarrow AD^{d-1} \qquad D = \mathrm{diag}\{\|a_1\|, \dots, \|a_N\|\}$$

- ↑ d emphasizes the significance of popular items,
- ↓ d improves sensitivity to rare/niche items,
- often leads to a significant quality improvement.

2. OBJECTIVES

To develop efficient LRA method that

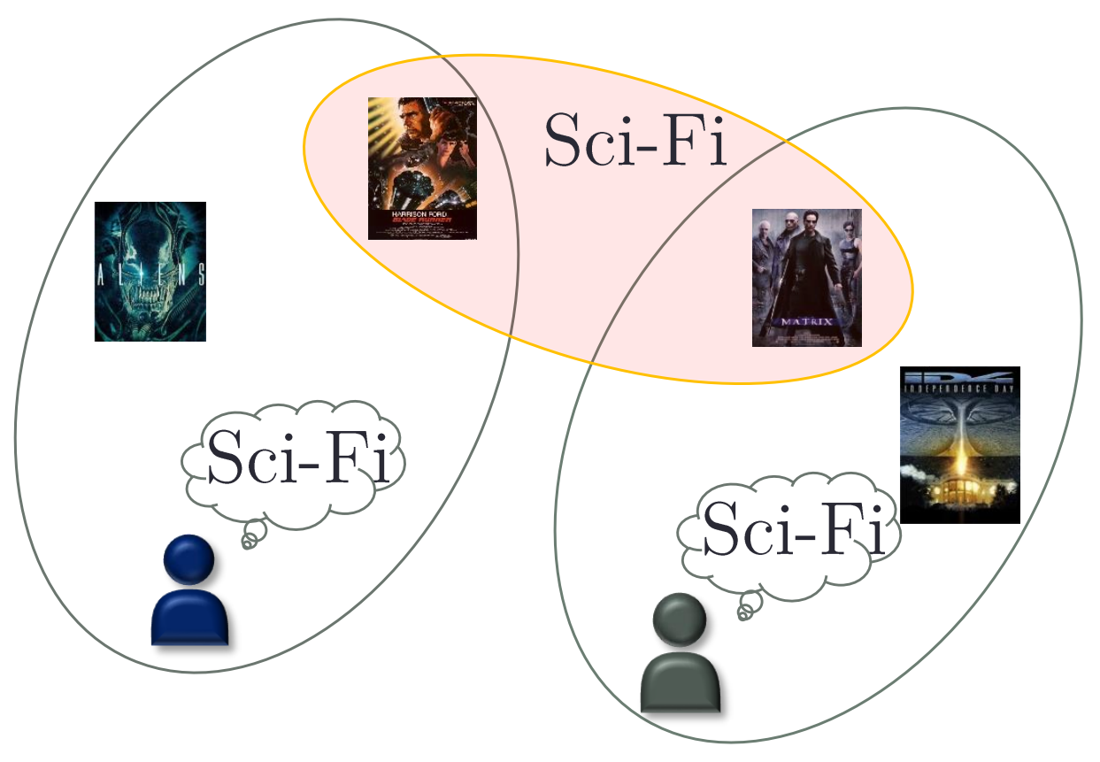
- allows incorporating side information into the model,
- inherits the key advantages of the PureSVD approach,
- admits minimal tuning,
- outperforms strong baselines in typical tasks.

3. KEY IDEA

"Similarity" of users i and j depends on co-occurrence of items in their preferences.

$$G = AA^\top = U\Sigma^2 U^\top \quad \leftrightarrow \quad g_{ij} = a_i^\top a_j \ \sim \mathrm{sim}(i,j)$$

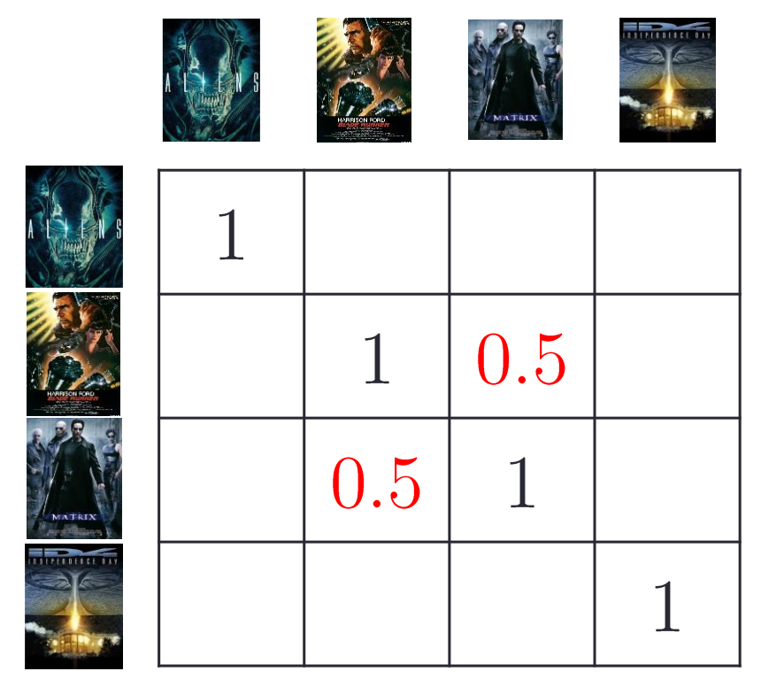
Replace scalar products with a bilinear form.



$$\mathrm{sim}(i,j) \sim a_i^\top S a_j$$

Creates "virtual" links based on side features.

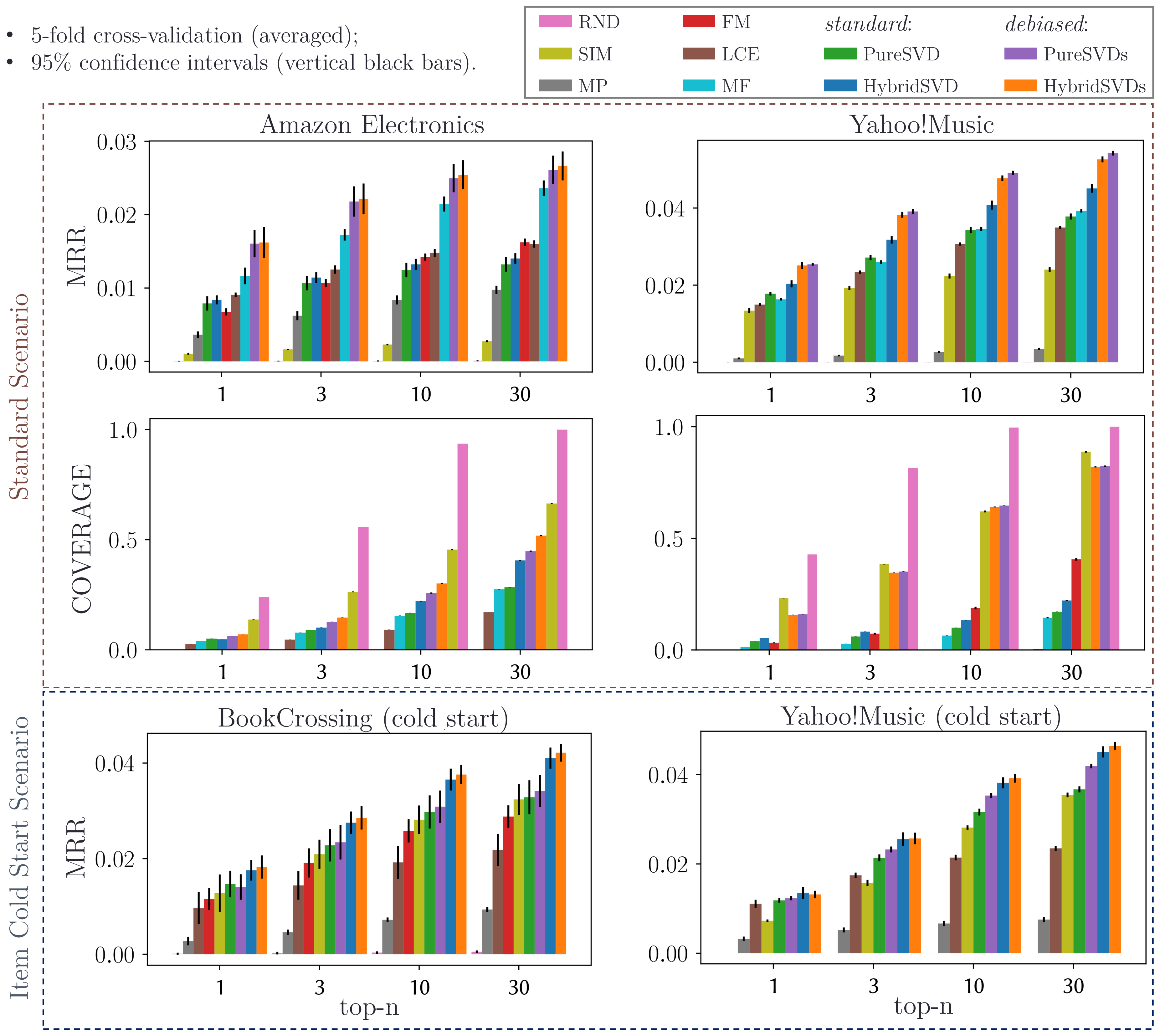
Similarity matrix S



1			
	1	0.5	
	0.5	1	
			1

7. RESULTS

- 5-fold cross-validation (averaged);
- 95% confidence intervals (vertical black bars).



Polara + binder = Reproducibility in browser!

Play with it on your own (no setup required), visit the link below for further instructions:
https://github.com/evfro/recsys19_hybridsvd

Polara – open-source recsys framework for quick and reproducible experimentation. (Note: I'm the author.)
https://github.com/evfro/polara

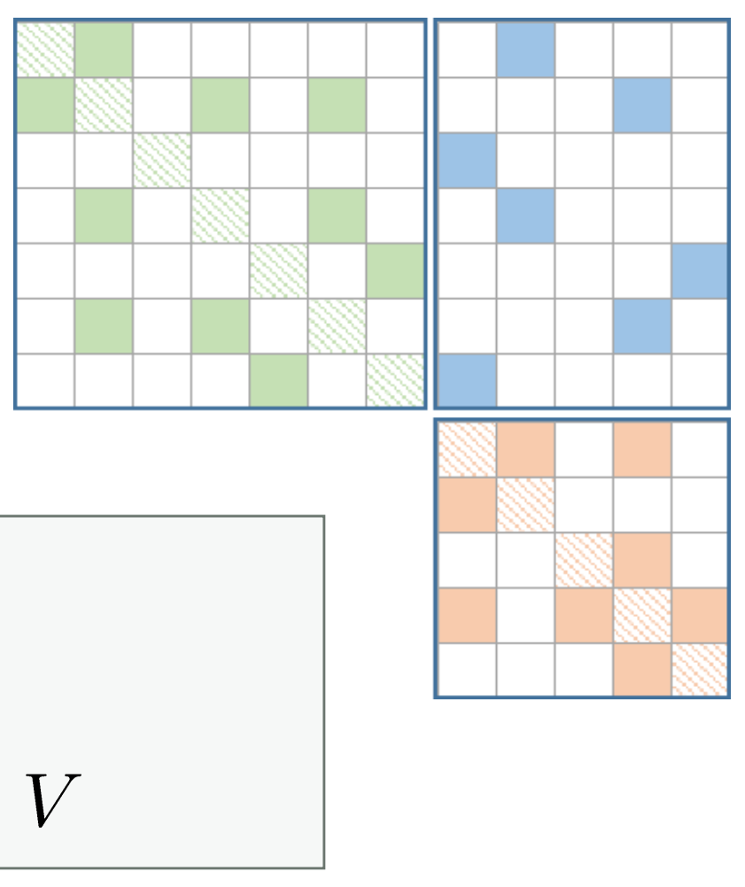
4. SOLUTION

The problem takes generalized form:

$$\begin{cases} AA^\top = U\Sigma^2 U^\top \\ A^\top A = V\Sigma^2 V^\top \end{cases} \Longrightarrow \begin{cases} ASA^\top = U\Sigma^2 U^\top \\ A^\top KA = V\Sigma^2 V^\top \end{cases}$$

Matrix "roots":

$$S = L_S L_S^\top \qquad K = L_K L_K^\top$$



Solution:

via SVD of an auxiliary matrix: [Abdi 2007; Allen/Grosenick/Taylor 2014] $L_K^\top A L_S = U\Sigma V^\top$

link to the original latent space: $L_K^{-\top} U = U, \qquad L_S^{-\top} V = V$

Properties:

latent space structure: $U^\top K U = I, \qquad V^\top S V = I$

"hybrid" folding-in: $p = L_S^{-T} V V^T L_S^T a$

+ everything from standard SVD

Controlling side features weight and similarity structure:

$$\begin{cases} S = (1-\alpha)I + \alpha Z, \\ K = (1-\beta)I + \beta W. \end{cases}$$

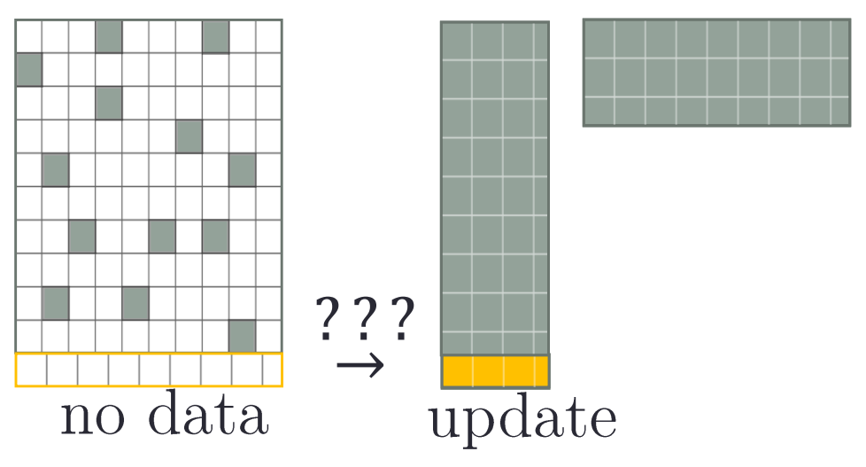
Z, W are real symmetric matrices, $-1 \le z_{ij}, w_{ij} \le 1 \ \forall\ i, j$;
$0 \le \alpha, \beta \le 1$ control how sensitive the model is to side features.
$\alpha = \beta = 0$ turns the model back into PureSVD.

5. ADDRESSING COLD START

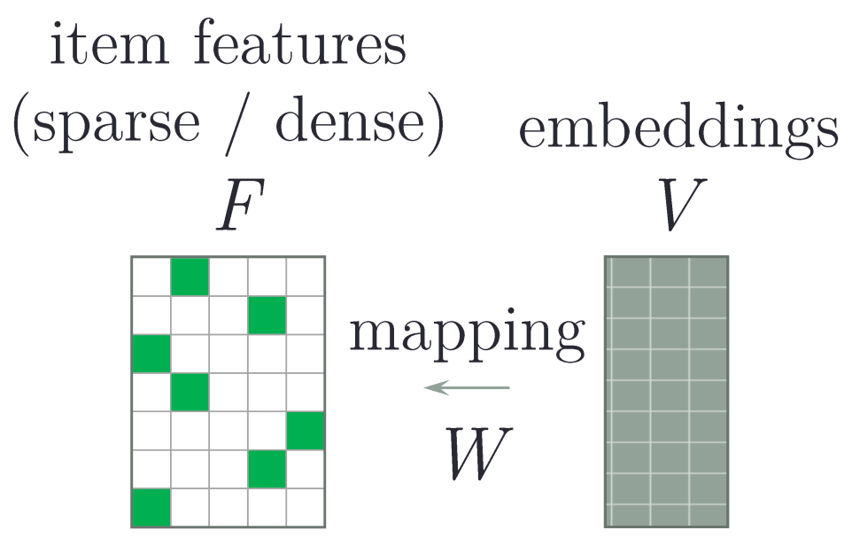
Finding a transformation between latent and real features is a hard task.
Hybrid MF models often incorporate this task into main optimization.

Several representative examples:

- Collective factorization [Singh/Gordon 2008, Acar/Kolda/Dunlavy 2011],
- Local Collective Embeddidngs (LCE) [Saveski/Mantrach 2014],
- Factorization Machines (FM) [Rendle 2010].



It is a lot simpler with SVD!



Using the S-orthogonality property:

$$VW = F \to W = V^\top SF$$ ← **analytic form**

Given any feature vector f, we can find the corresponding embedding v :
$$W^\top v = f$$ ← **quick to solve**

Predicted relevance scores:
$$p = U\Sigma v = AVv$$

Works for PureSVD! (set $S = I,\ K = I$)

6. EFFICIENT COMPUTATIONS

Overall computational complexity:

$$\underbrace{\mathrm{O}(\mathrm{nnz}_A \cdot r) + \mathrm{O}((M+N)\cdot r^2)}_{\text{PureSVD part}} + \underbrace{\mathrm{O}((J_K + J_S)\cdot r))}_{\text{“hybrid” part}}$$

"matvec" complexities for L_K and L_S

Finding Cholesky/square root factors:

- efficient schemes for sparse and dense representations;
- computational complexity is adjustable, depends on data structure;
- can be computed on GPU https://developer.nvidia.com/cholmod.

For sparse feature representations:

- symbolic Cholesky decomposition, can reuse sparsity pattern for different α and β.
- can also use incomplete or thresholded variant of Cholesky decomposition

For dense feature representations:

- fast symmetric factorization for computing matrix square root [Ambikasaran/O'Neil/Singh 2014].

8. CONCLUSIONS

HybridSVD offers a set of practical advantages:

- Allows generating structured latent feature space.
- Has a small number of intuitive hyper-parameters.
- Enables simplified tuning.
- Supports quick online and session-based recommendations.
- Effective in standard, warm start, and cold start regimes.

May not be the best in all cases; however, definitely is a strong baseline!

9. SOME REFERENCES

Improving PureSVD with EigenRec model:
Nikolakopoulos, A. N., Kalantzis, V., Gallopoulos, E., & Garofalakis, J. D. (2019). EigenRec: generalizing PureSVD for effective and efficient top-N recommendations. *Knowledge and Information Systems*, 58(1), 59-81.

A structural view on generalized SVD:
Allen, G. I., Grosenick, L., & Taylor, J. (2014). A generalized least-square matrix decomposition. *Journal of the American Statistical Association*, 109(505), 145-159.