HybridSVD

When Collaborative Information is Not Enough

Evgeny Frolov, Ivan Oseledets

State of the Art: revisited

Proper tuning of baseline models vs. modern SOTA shows surprising results.

- DL vs. MF in standard CF task, J. Basilico talk at ICML 2019
- In session-based recommendations [Ludewig/Jannach 2018]
- On the difficulty of evaluating baselines [Rendle/Zhang/Koren 2019]
- In other disciplines, e.g., NLP [Levy/Goldberg/Dagan 2015]

Are we really making much progress? [Dacrema/Cremonesi/Jannach 2019]

Best paper award

+ "hot" discussion in twitter.

Part of the reason – tuning can be hard and cumbersome.

Tuning is an Art

Grid search is a non-trivial problem.

Many techniques exist, e.g., Bayesian estimators, TPE, spectral approach.

Complexity **grows exponentially** with the number of hyper-parameters.

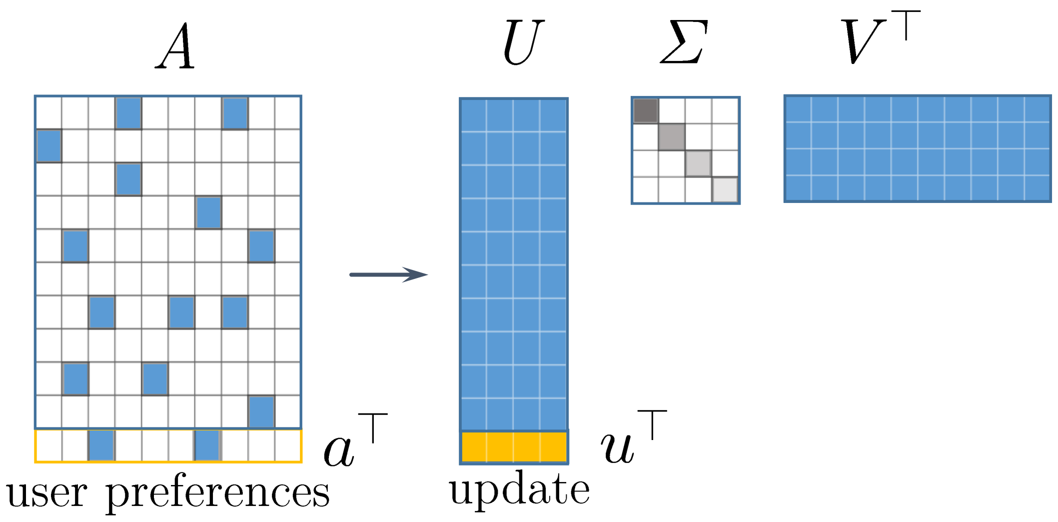
Example: tuning the number of latent features (rank) in MF.

Have to consider:

- wide range of values,
- varying training time,
- interdependencies with other hyper-parameters,
  like regularization, number of iterations, etc.

Can we reduce tuning hassles at least for the baselines?

Quick recap on SVD



PureSVD model, [Cremonesi/Koren/Turrin 2010], unknowns are replaced with zeros in A:

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

on-the-fly predictions (encoder-decoder like):

$$p = VV^\top a$$

predicted item scores (any) user preferences

valid for both existing and new users (folding-in)

Advantages:

- ✓ simple tuning via rank truncation (train once!)
- ✓ minimal storage requirements
- ✓ supports online and session-based predictions
- ✓ stable, deterministic output
- ✓ highly optimized implementations
- ✓ scales to ~billion-size problems, e.g., https://github.com/criteo/Spark-RSVD.

PureSVD vs. Weighted Matrix Factorization

Recommendations' quality of the most popular collaborative filtering techniques on

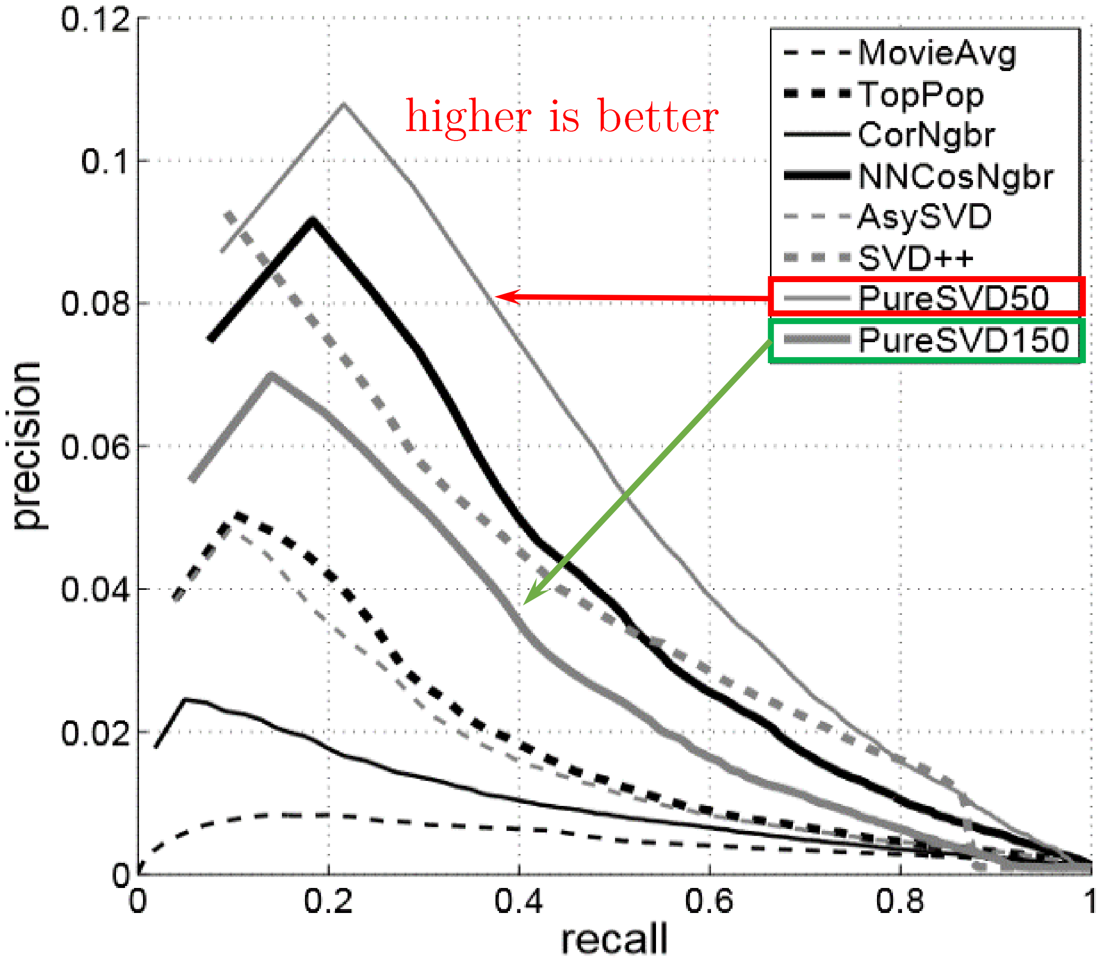
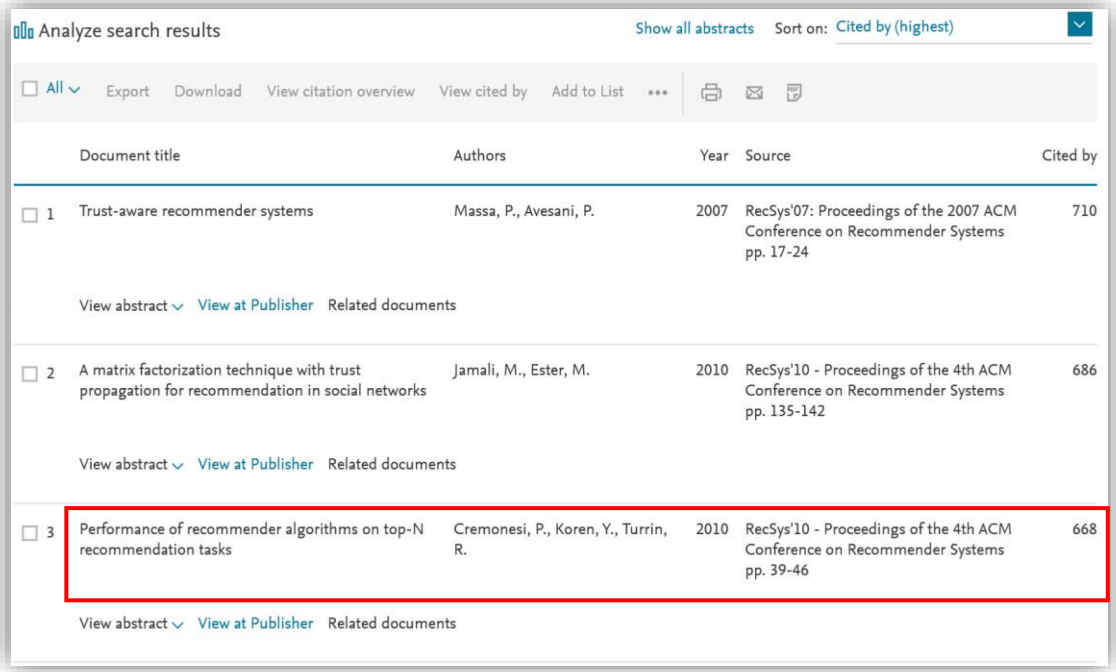


Image from: [Cremonesi/Koren/Turrin 2010].

Remarks:

- we are solving a surrogate problem (not a ranking problem)
- not (strictly) a matrix completion
- yet good enough for top-n recommendations

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



Improving SVD further

Input data balancing exists in other MF approaches.

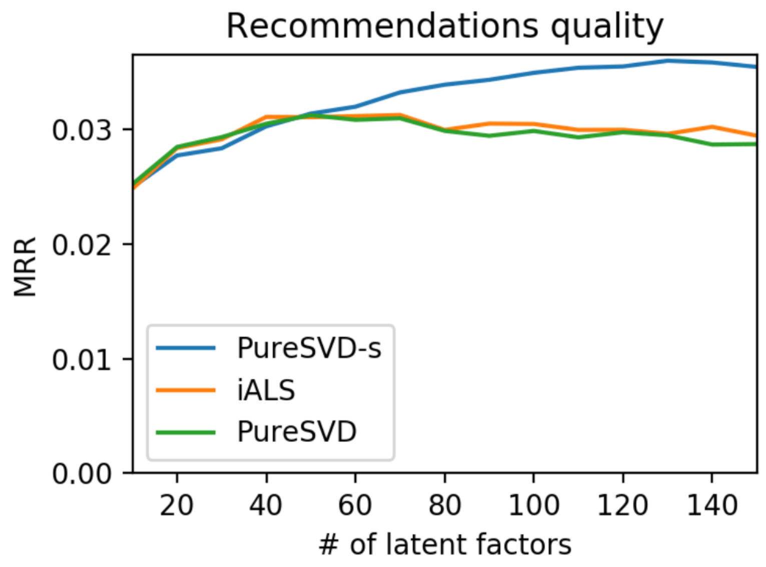
- ALS-based: Weighted Regularized Matrix Factorization (e.g., iALS [Hu/Koren/Volinsky 2008]);
- SGD-based: implicitly via negative sampling + custom objectives.

Simple data-debiasing trick for PureSVD:

from EigenRec model, [Nikolakopoulos et al. 2019]

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

- $\uparrow d$ emphasizes the significance of popular items;
- $\downarrow d$ improves sensitivity to rare/niche items;
- often increases diversity along with accuracy;
- optimal values of d typically lay in (0, 1) interval.



Experiment on Movielens-10M data, more details at: http://eigentheories.com/blog/to-svd-or-not-to-svd/

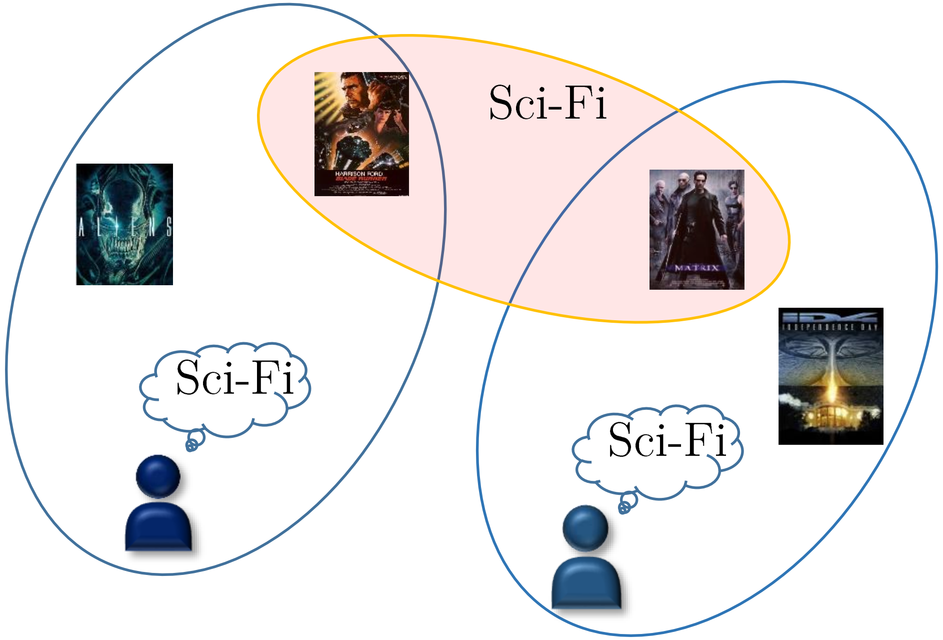
“Hybridization” of PureSVD

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

$$G = AA^\top = U\Sigma^2U^\top \quad \leftrightarrow \quad g_{ij} = a_i^\top a_j$$

Key idea: 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

HybridSVD model

$$\begin{cases} AA^\top = U\Sigma^2U^\top \\ A^\top A = V\Sigma^2V^\top \end{cases} \longrightarrow \begin{cases} A{\color{red}S}A^\top = U\Sigma^2U^\top \\ A^\top {\color{red}K} A = V\Sigma^2V^\top \end{cases}$$

Matrix "roots":

$$K = L_K L_K^\top, \; S = L_S L_S^\top$$

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:

"hybrid" folding-in:

$$p = L_S^{-T} V V^T L_S^T a$$

latent space structure:

$$U^\top K U = I, \qquad V^\top S V = I$$

+ everything from standard SVD (e.g., simplified rank tuning)

HybridSVD parameters

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

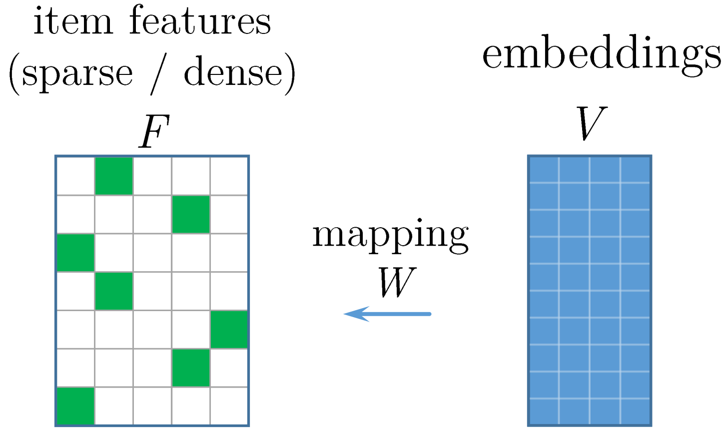
$0 \leq \alpha, \beta \leq 1$ control how sensitive the model is to side features.

Z, W are real symmetric matrices, $-1 \leq z_{ij}, w_{ij} \leq 1 \ \ \forall\ i, j$.

- single hyper-parameter per entity type (user / item),
- straightforward impact on model behavior.

- freedom to choose similarity measure;
- efficient computations for different feature representations:
  - sparse Cholesky for sparse features,
  - fast symmetric factorization ([Ambikasaran et al. 2014]) for dense features;
- overall computational complexity is adjustable, data-dependent.

Addressing cold start with SVD



We hope that special structure of the HybridSVD's latent space will make it easier to recover a "good" mapping.

Feature mapping computation is decoupled from training:

$$VW = F \rightarrow W = V^\top SF$$ ← **analytic solution**

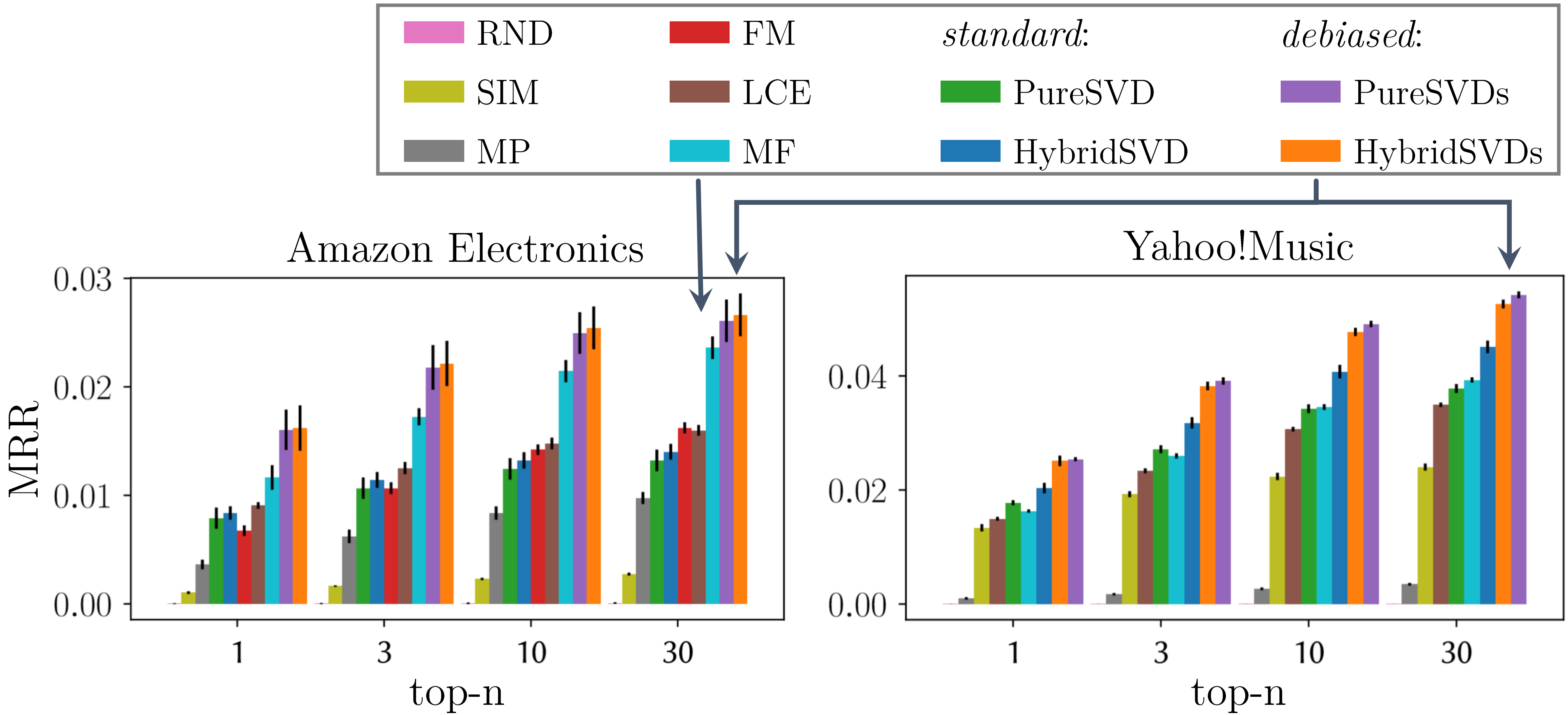
Given any feature vector f, we find the corresponding embedding v from:
$$W^\top v = f$$
← **quick to solve**

Relevance scores prediction:

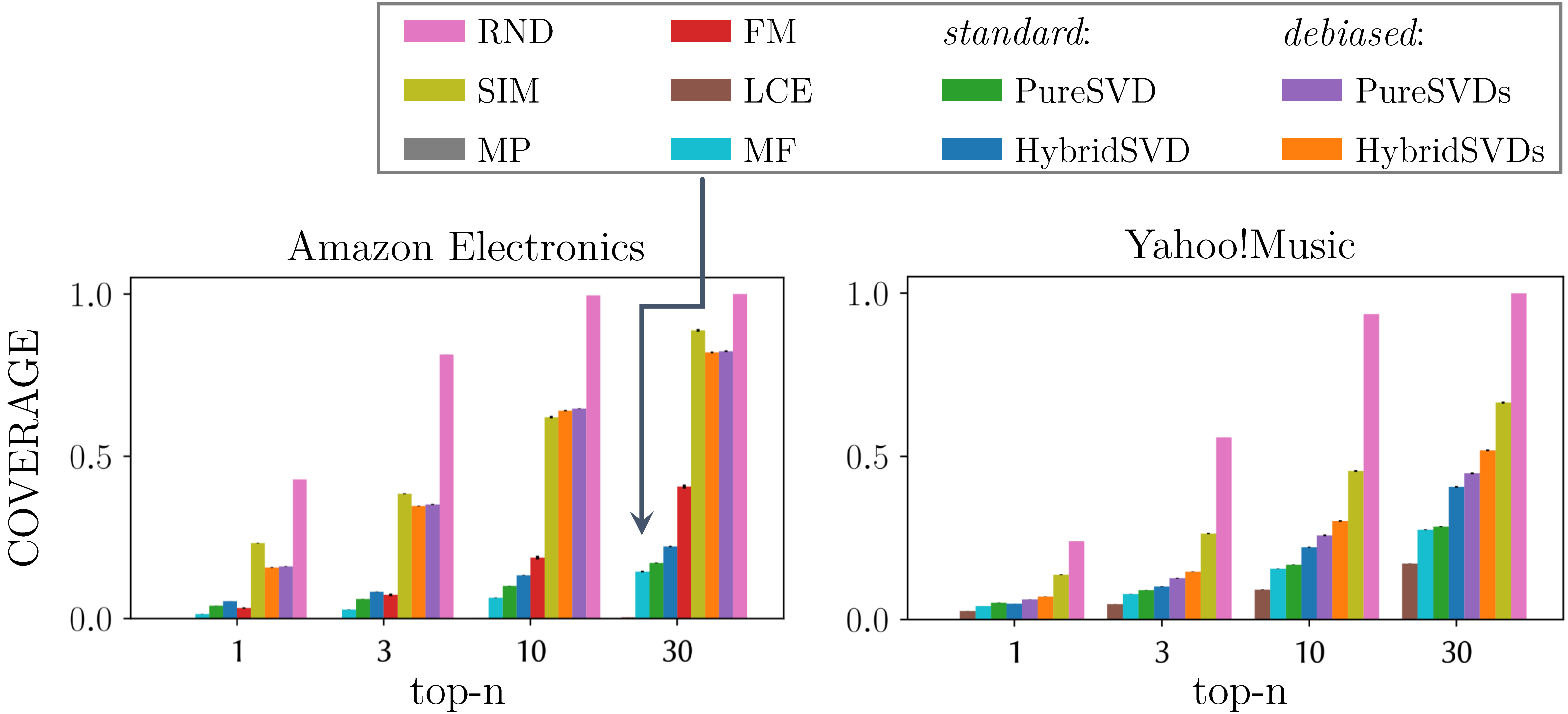
$$p = U\Sigma v = AVv$$

Works for PureSVD as well by setting $S = I$ and $K = I$.

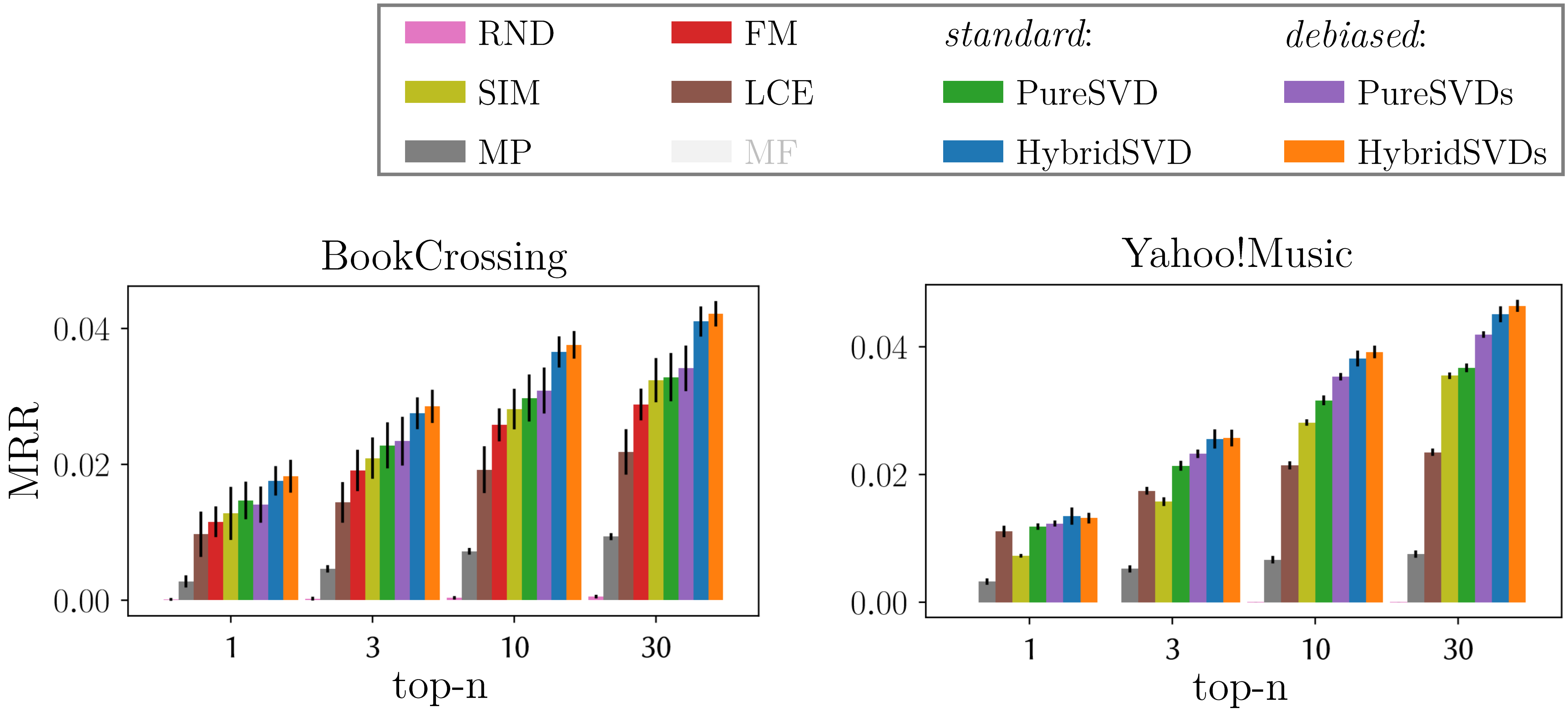
Evaluation in standard scenario



Evaluation in standard scenario



Evaluation in cold start scenario



Surprisingly, in some cases even PureSVD performs better then more sophisticated hybrid models.

Also see comparison with LightFM at https://www.eigentheories.com/blog/lightfm-vs-hybridsvd/.

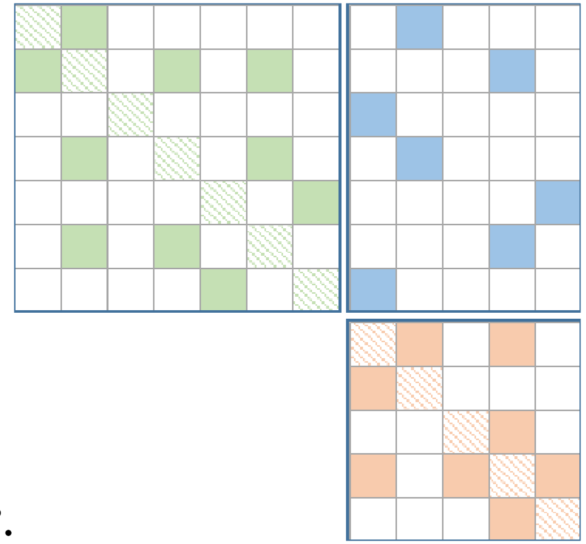
Polara + binder = Reproducibility in browser

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. Disclaimer: I'm the author.
https://github.com/evfro/polara

More examples on reproducing others work (and not only) can be found in Polara repository.

Conclusions



HybridSVD is simple, efficient and very competitive.

- ✓ Allows generating structured latent feature space.
- ✓ Has a *small number* of hyper-parameters with *intuitive effects*.
- ✓ Enables quick tuning on a grid via rank truncation.
- ✓ Supports dynamic 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**!

- ❑ Requires a bit more work at the data preprocessing step (Cholesky|Square root).
- ❑ In the case of non-binary rating data may lead to spurious correlations, fixed by tensor formulation, see [Frolov/Oseledets 2018] (work in progress).

Some references

- HybridSVD paper:
  Frolov E, Oseledets I. *HybridSVD: when collaborative information is not enough.* InProceedings of the 13th ACM Conference on Recommender Systems 2019 Sep 10 (pp. 331-339). ACM.
- Improving PureSVD results 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.
- Fast matrix square root computation for "identity + low rank" matrices:
  Ambikasaran, S., O'Neil, M., & Singh, K. R. (2014). *Fast symmetric factorization of hierarchical matrices with applications.* arXiv preprint arXiv:1405.0223.
- 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.

Thank you!

evgeny.frolov@skoltech.ru

https://www.eigentheories.com