HybridSVD
When Collaborative Information is Not Enough

Evgeny Frolov, Ivan Oseledets

September 18, 2019
Copenhagen, Denmark
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 $A$:

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

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.

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)

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]

\[ A \leftarrow AD^{d-1}, \quad D = \text{diag}\{\|a_1\|, \ldots, \|a_N\|\} \]

• ↑ \( d \) emphasizes the significance of popular items;
• ↓ \( d \) 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.

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

**Key idea:** replace scalar products with a bilinear form.

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

Creates “virtual” links based on side features.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
HybridSVD model

\[
\begin{align*}
AA^\top &= U \Sigma^2 U^\top \\
A^\top A &= V \Sigma^2 V^\top
\end{align*}
\]

\[
\begin{align*}
AS A^\top &= U \Sigma^2 U^\top \\
A^\top KA &= V \Sigma^2 V^\top
\end{align*}
\]

Matrix “roots”:
\[
K = L_K L_K^\top, \quad S = L_S L_S^\top
\]

Solution:

via SVD of an auxiliary matrix:
[Abdi 2007; Allen/Grosenick/Taylor 2014]

\[
L_K^\top AL_S = U \Sigma V^\top
\]

link to the original latent space:
\[
L_K^{-\top} U = U, \quad L_S^{-\top} V = V
\]

Properties:

“hybrid” folding-in:
\[
p = L_S^{-\top} V V^\top L_S^\top a
\]

latent space structure:
\[
U^\top K U = I, \quad V^\top S V = I
\]

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

\[
\begin{align*}
S &= (1 - \alpha)I + \alpha Z, \\
K &= (1 - \beta)I + \beta W.
\end{align*}
\]

0 ≤ α, β ≤ 1 control how sensitive the model is to side features.

Z, W are real symmetric matrices, -1 ≤ z_{ij}, w_{ij} ≤ 1 ∀ i, j.

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

- freedom to choose similarity measure;
- efficient computations for sparse/dense representation;
- overall computational complexity is adjustable, data-dependent.
Addressing cold start with SVD

Feature mapping computation is decoupled from training:

\[ VW = F \rightarrow W = V^\top SF \]

\[ W^\top v = f \]

Given any feature vector \( f \), we find the corresponding embedding \( v \) from:

\[ p = U\Sigma v = AVv \]

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

Works for PureSVD as well by setting \( S = I \) and \( K = I \).
Evaluation in standard scenario

Amazon Electronics

Yahoo!Music

MRR

top-n

1 3 10 30

RND FM SIM LCE MP MF

standard: PureSVD PureSVDs

debiased: HybridSVD HybridSVDs
Evaluation in standard scenario

Amazon Electronics

Yahoo!Music

coverage vs. top-n
Evaluation in cold start scenario

<table>
<thead>
<tr>
<th>RND</th>
<th>FM</th>
<th>SIM</th>
<th>LCE</th>
<th>MP</th>
<th>MF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

standard: PureSVD  debiased: PureSVDs HybridSVD HybridSVDs

BookCrossing

Yahoo!Music

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

• Improving PureSVD results with EigenRec model

• A simple generalized SVD solution:

• A structural view on generalized SVD:

Thank you!

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