Skoltech

Skolkovo Institute of Science and Technology

HybridSVD for feature selection

Introduction to Recommender Systems

Mekan Hojayev

Abdul Aziz Samra

Outline

1. Introduction
2. Problem Statement
3. Why PureSVD is not relevant?
4. Hypothesis and experiments
5. Conclusions

PureSVD is equivalent to an eigenproblem for the scaled cosine similarity matrix

$$\mathrm{sim}(i,j) \sim \bar{a}_i^\top \bar{a}_j \qquad \mathrm{sim}(i,j) \sim \bar{a}_i^\top S\, \bar{a}_j$$

a_i is an i^{th} column of matrix A

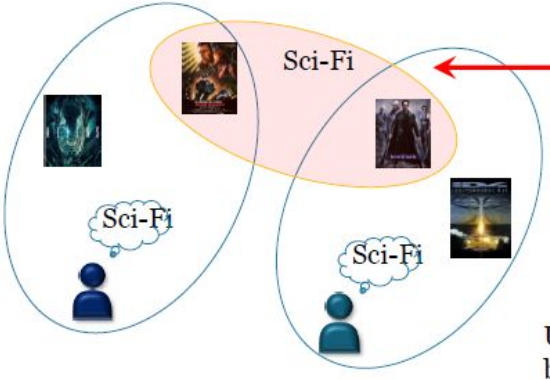
Off-diagonal elements of S encode a similarity between items based on their features.

Key idea: substitute scalar products with a bilinear form that takes side information into account.

“similarity” of users i and j depends on co occurrence of items in their preferences

$$C = A^{\top}A = V\Sigma^2V^{\top} \quad \leftrightarrow \quad c_{ij} = \bar{\boldsymbol{a}}_i^{\top}\bar{\boldsymbol{a}}_j$$

new formulation: $c_{ij} \sim \bar{\boldsymbol{a}}_i^{\top} S\, \bar{\boldsymbol{a}}_j$



“Virtual” connection based on item features.

User-based “side” similarity can be exploited in similar fashion.

Similarity matrix S

1			
	1	0.5	
	0.5	1	
			1

1. Build SPD similarity matrices K, S for users and items based on *side information*.
2. Solve a *new eigendecomposition* problem:

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

Σ is a diagonal matrix of singular values.

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

Solution:

via SVD of an auxiliary matrix [Abdi 2007; Allen et al. 2014]: $L_K^\top A L_S = \hat{U}\Sigma\hat{V}^\top, \quad L_K L_K^\top = K, \; L_S L_S^\top = S$

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

Properties:

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

"hybrid" folding-in: $\boldsymbol{p} = L_S^{-\top}\hat{V}\hat{V}^\top L_S^\top \boldsymbol{a}.$

HybridSVD parameters

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

- Z, W are real symmetric matrices

$$-1 \leq z_{ij}, w_{ij} \leq 1 \ \forall\, i, j$$

- $0 \leq \alpha, \beta \leq 1$ are free model parameters

control contribution of features into the model.

- Max 3 hyperparameters to tune: α, β and the rank of decomposition,
- Simplified procedure with rank truncation for any fixed pair of α and β.

Cold start with SVD

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

Z and W are users/items similarity in the features space.

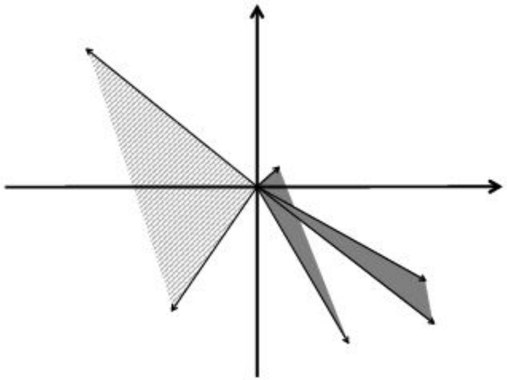
how to select the best features to include while calculating these matrices?
in other words, **introduce relevant ranking scheme for features**

Hypothesis

find a **compressed version of a feature matrix F** (which was used to calculate S) enriched with the correlations data computed based on ratings matrix A, then apply *maxvol algorithm* to select the most important features.

MAXVOL (Maximal-Volume Algorithm)

…algorithm, searches very efficiently for a submatrix of a factor matrix with the locally maximal determinant.



. An illustration of the intuition behind Maxvol for searching the seed

Related work: https://arxiv.org/pdf/1610.04850.pdf

Hypothesis

compressed version of a feature matrix F

Hypothesis 1

$$S_A = f(A^T A) = L_{S_A} L_{S_A}^T$$

$$L_{S_A}^T F_i = U \Sigma V^T$$

$$\text{feature ranking} = \text{maxvol}\,(V)$$

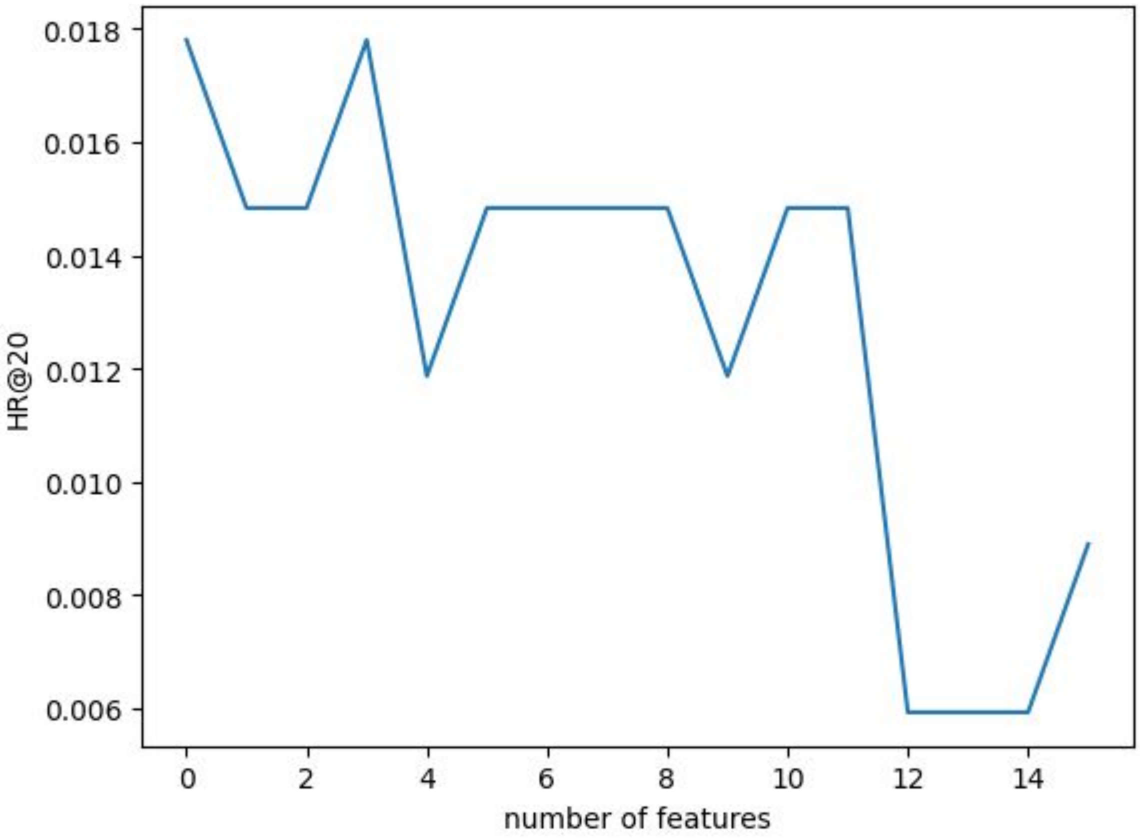
Hypothesis 2

$$B = AF_i$$

$$S_F = f(B^T B) = L_{S_F} L_{S_F}^T$$

$$F_i L_{S_F} = U \Sigma V^T$$

$$\text{feature ranking} = \text{maxvol}\,(V)$$


0.018
0.016
0.014
0.012
0.010
0.008
0.006
HR@20
0
2
4
6
8
10
12
14
number of features

Thank you for attention!