Skoltech

Skolkovo Institute of Science and Technology

Factorization in recommender systems

Nikolai Goncharov

Denis Rollov

Svetlana Gabdullina

2020

Skoltech
Skolkovo Institute of Science and Technology

Introduction

Applications of the SVD:

- Solving linear equations
- Least squares minimization
- Low-rank matrix approximation
- Signal processing
- Recommender systems

Skoltech
Skolkovo Institute of Science and Technology

Introduction

Applications of the SVD:

- Solving linear equations
- Least squares minimization
- Low-rank matrix approximation
- Signal processing
- **Recommender systems**

Skoltech
Skolkovo Institute of Science and Technology

SVD for 2D recommender systems

A: User × Item

SVD: $A = U\Sigma V^\top = U\Sigma^{1/2}\Sigma^{1/2}V^\top = PQ^\top$

- U for user latent factors
- Σ for the strength of each latent factor
- V for item latent factors

Prediction of rating[1]: $r_{ij} = p_i q_j^\top$

Skoltech
Skolkovo Institute of Science and Technology

How we can extend it?

Add more information and work with tensors:

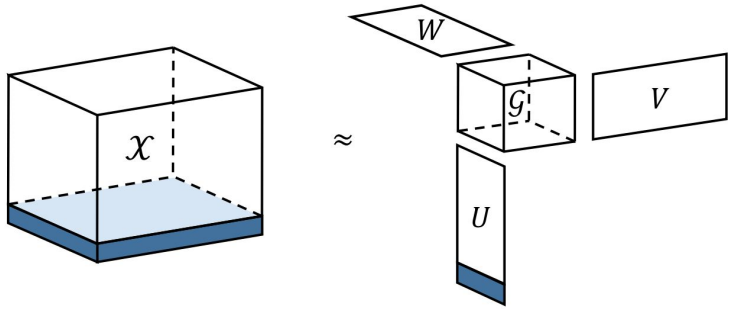
- **feedback** - rating from the particular user for the particular film
  or
- **context** - month when the particular user watched the particular film

Skoltech
Skolkovo Institute of Science and Technology

Tensor decomposition

Tucker decomposition: $X \approx G \times_1 U \times_2 V \times_3 W$

- U for user
- V for film
- W for feedback or context



HOOI method implemented in [tensorly.decomposition.tucker]

Prediction of rating[2]: $R_{(i)} = VV^\top P_{(i)} WW^\top$

- $P_{(i)}$ is a binary matrix of an i-th user's preferences
- $R_{(i)}$ is a matrix of recommendations

Skoltech
Skolkovo Institute of Science and Technology

Data preprocessing

Dataset Movielens

Tensor 1: user_id × movie_id × rating

Tensor 2: user_id × movie_id × month (when rating was assigned)

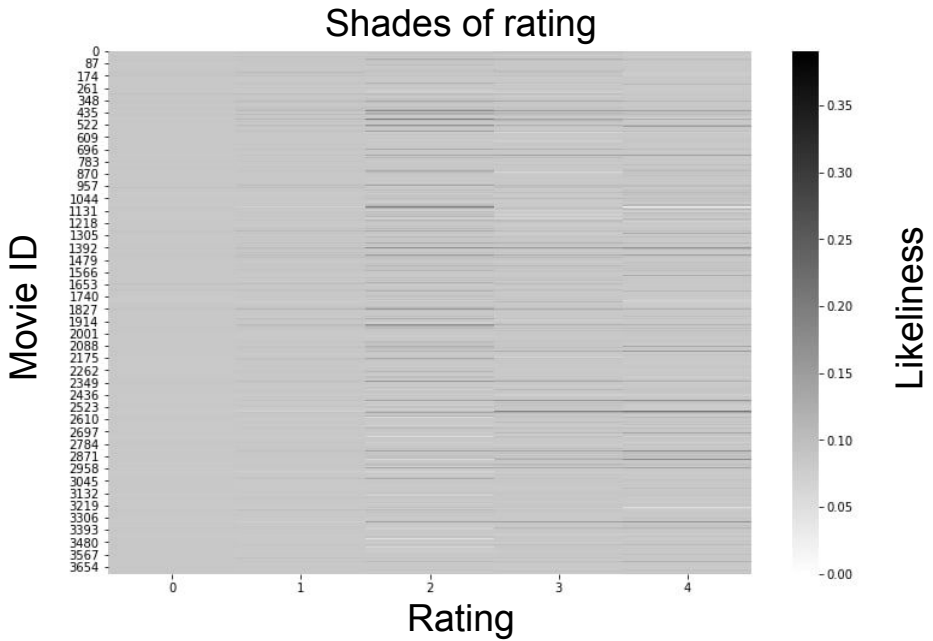
Train-Test splitting:

1. map tensor axis indices with corresponding ids
2. random shuffle the tensor by user_id
3. split by user_id, 70% train and 30% test

Skoltech
Skolkovo Institute of Science and Technology

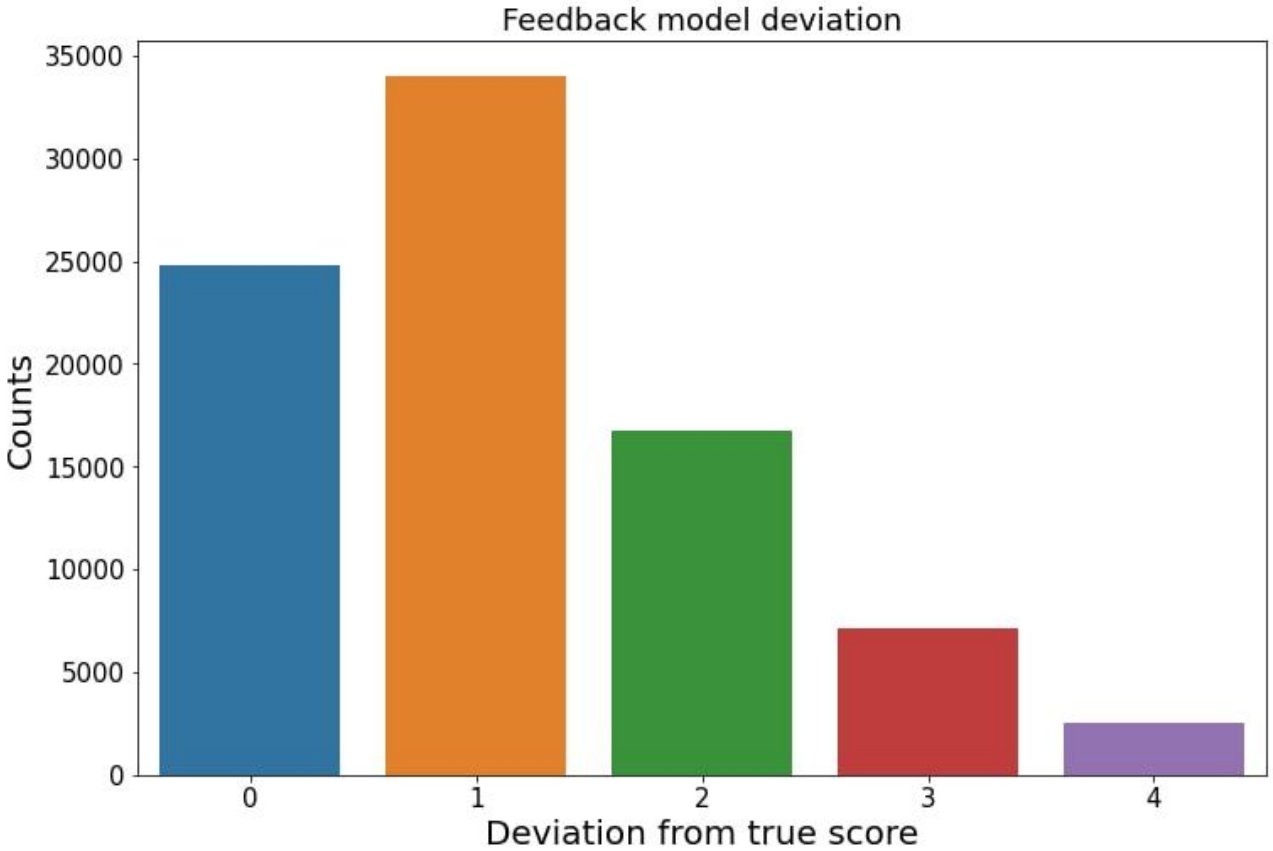
Recommend process: feedback model

1. Delete 30% of units in a slice for i-th test user
2. Use formula of recommendation $R_{(i)} = VV^\top P_{(i)} WW^\top$
3. Get the result



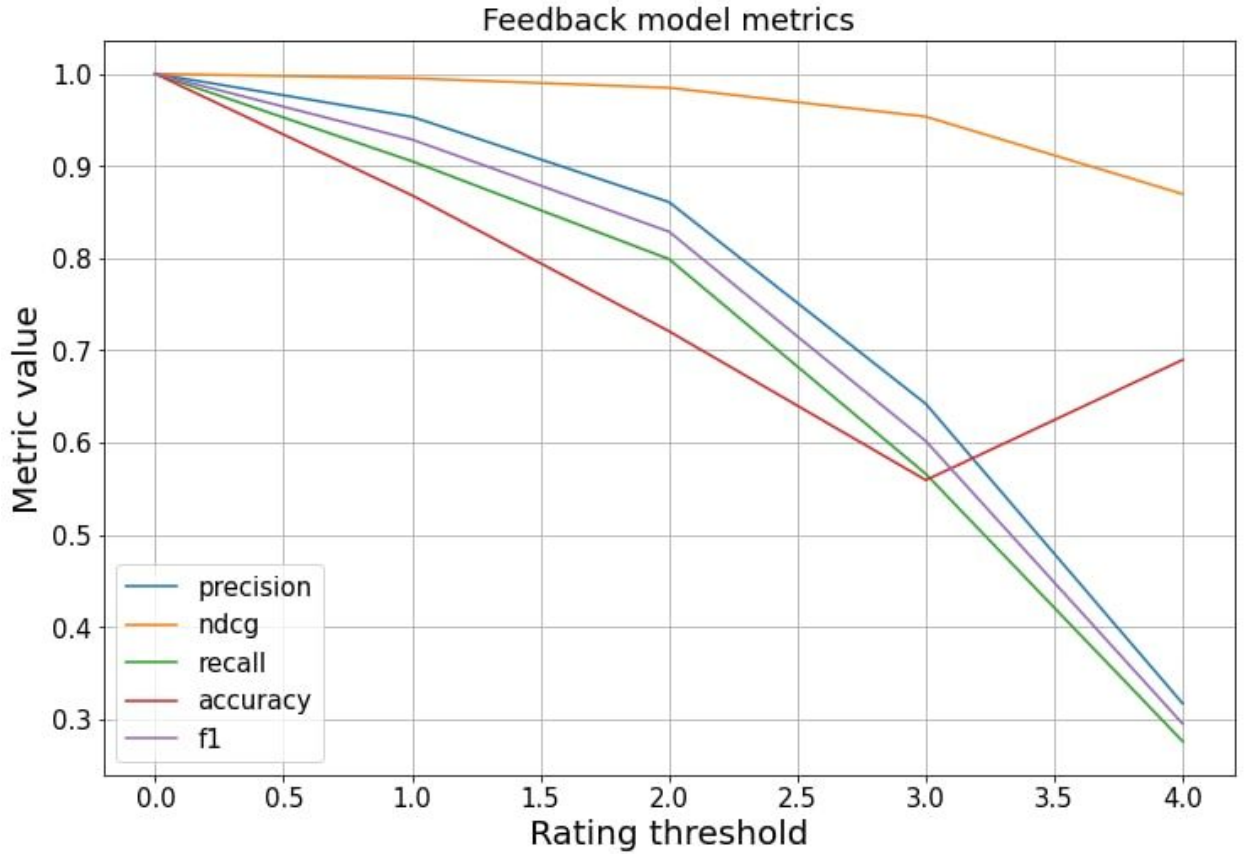
Skoltech
Skolkovo Institute of Science and Technology

Validation: feedback model



Skoltech
Skolkovo Institute of Science and Technology

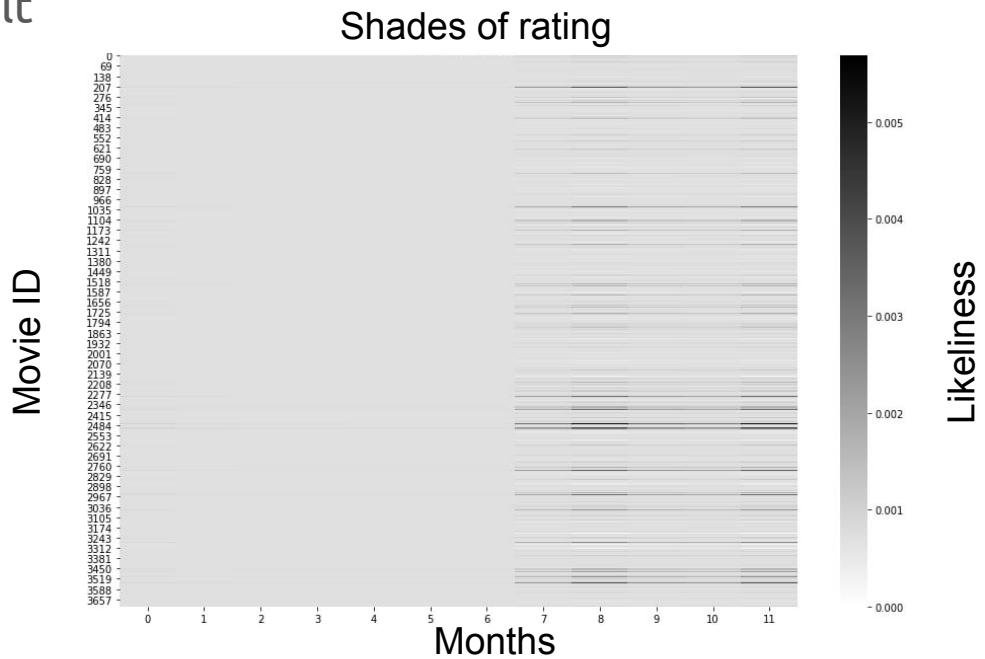
Validation: feedback model



Skoltech
Skolkovo Institute of Science and Technology

Recommend process: context model

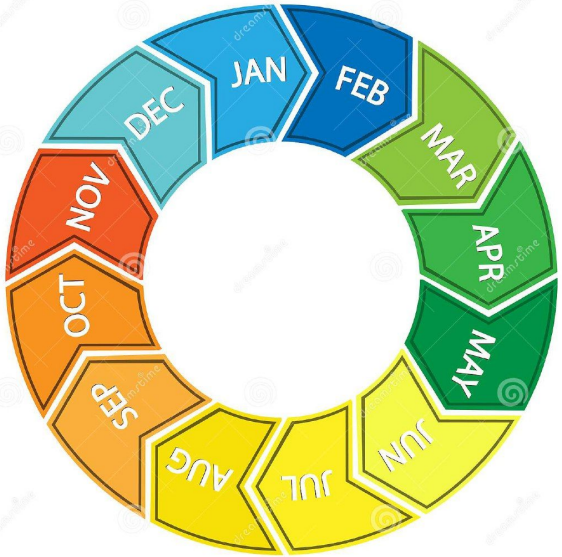
1. Delete 30% of units in a slice for i-th test user
2. Use formula of recommendation
3. Get the result



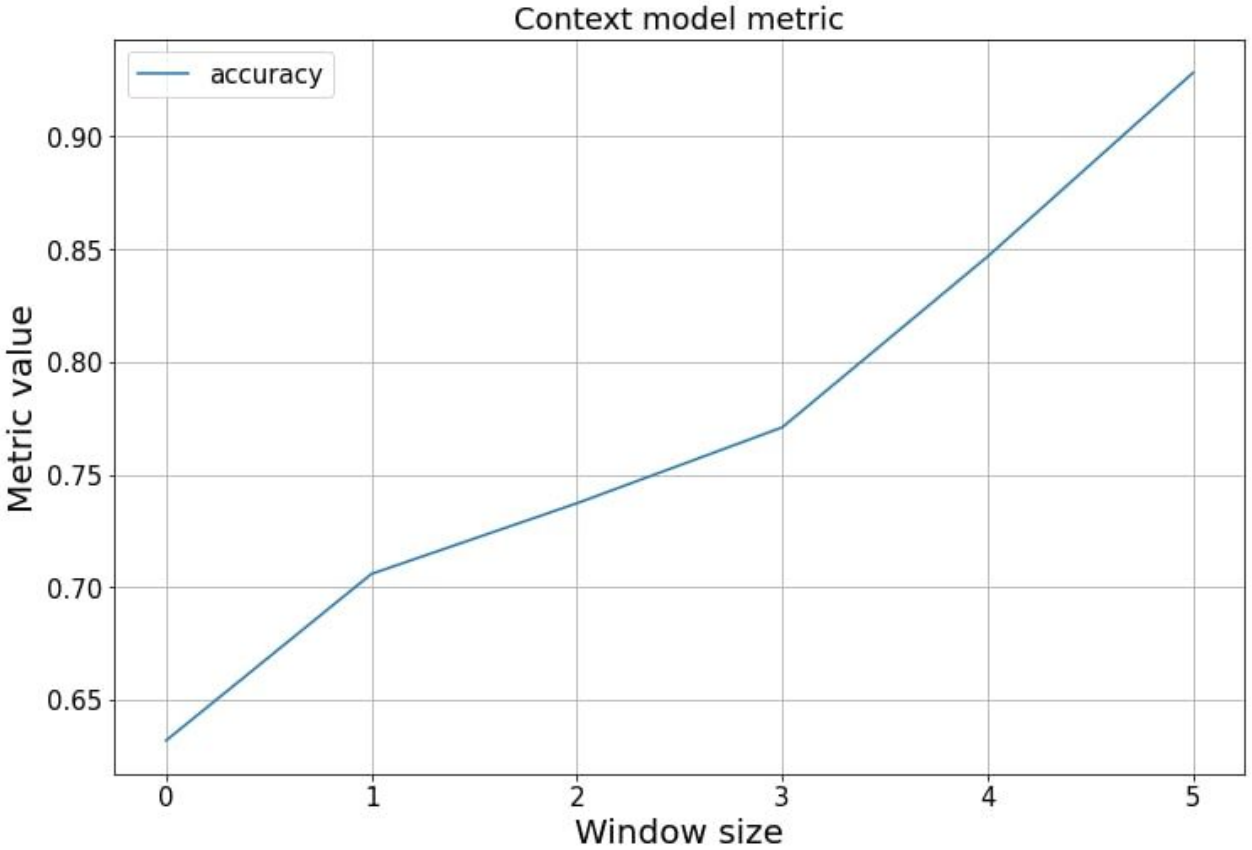
Skoltech
Skolkovo Institute of Science and Technology

Validation: context model

1. Choose parameter w
2. $$m_{pred} = \begin{cases} 1, & \text{if } m_{pred} \in [m_{true} - w; m_{true} + w] \\ 0, & \text{otherwise} \end{cases}$$
3. $$accuracy = \frac{\sum_n m_{pred,i}}{N}$$



Validation: context model



Skoltech
Skolkovo Institute of Science and Technology

Demo recommendations

Test user's favorite movies

	name
0	Toy Story (1995)
33	Babe (1995)
436	Dave (1993)
476	Jurassic Park (1993)
496	Mrs. Doubtfire (1993)
583	Ghost (1990)
1000	Parent Trap, The (1961)
1179	Princess Bride, The (1987)
1271	Indiana Jones and the Last Crusade (1989)
2016	101 Dalmatians (1961)
2692	Iron Giant, The (1999)

We recommend

	name
1	Jumanji (1995)
642	Mission: Impossible (1996)
1219	Local Hero (1983)
1236	Duck Soup (1933)
1263	High Noon (1952)
1899	Breakfast Club, The (1985)
2011	Lady and the Tramp (1955)
2031	Splash (1984)
2124	Willow (1988)
2719	And Now for Something Completely Different (1971)
2728	Big (1988)
3352	Animal House (1978)

Skoltech
Skolkovo Institute of Science and Technology

Sources

- Matrix factorization techniques for recommender systems, Yehuda Koren, Robert Bell, Chris Volinsky, 2009
- Fifty Shades of Ratings, Evgeny Frolov, Ivan Oseledets, 2016
- Tensor Methods and Recommender Systems, Evgeny Frolov, Ivan Oseledets, 2016
- Tensor Decompositions and Applications, Tamara G. Kolda, Brett W. Bader 2009,