SVD RS II (Optional)

STAT3009 Recommender Systems

by Ben Dai (CUHK)
on Department of Statistics and Data Science

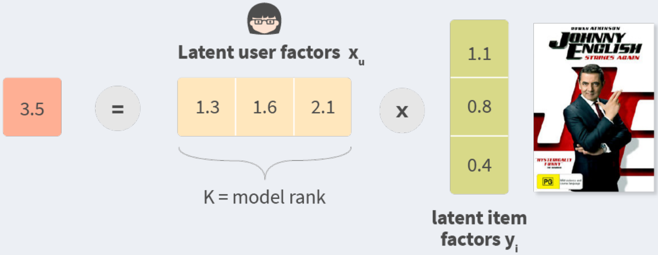
» Matrix Factorization (MF) Models

* SVD-based collaborative filtering
* Alternating Least Squares (ALS) optimization
* Coordinate descent methods
* Low-rank matrix completion

» Recall: Matrix factorization

The main idea of *matrix factorization*

- *params* $\rightarrow$ users/items preference
- *inner-production* $\rightarrow$ user-item interaction
- (first proposed by **Simon Funk** during the Netflix Prize)



The idea of MF RS (Source[1]). The latent factors $\mathbf{x}_u$ and $\mathbf{y}_i$ in the image are denoted as $\mathbf{p}_u$ and $\mathbf{q}_i$ in the slides.

---

[1] Image Source: https://blog.griddynamics.com/how-deep-learning-improves-recommendations-for-your-catalog/

» Recall: Matrix Factorization

Step 1 Introduce a method with some *parameters* (latent factors)

- Model the user-item interaction as the *inner product*

$$\hat{r}_{ui} = \boldsymbol{p}_u^\intercal \boldsymbol{q}_i \to r_{ui}$$

Step 2 Estimate the *parameters* by minimizing the regularized *MSE*

$$\min_{\boldsymbol{P},\boldsymbol{Q}} \frac{1}{|\Omega|} \sum_{(u,i)\in\Omega} (r_{ui} - \boldsymbol{p}_u^\intercal \boldsymbol{q}_i)^2 + \lambda \Big(\sum_{u=1}^{n} \|\boldsymbol{p}_u\|_2^2 + \sum_{i=1}^{m} \|\boldsymbol{q}_i\|_2^2 \Big) \quad (1)$$

- Use *cross-validation* to determine the optimal *hyperparameters* (K, λ), denoted as K^* and λ^*
- Refit the model based on the *full training data* with K^* and λ^*.

Step 3 Use the *estimated model* with the *best hyperparameters* to make predictions

» Design Your Own Recommender System

We also summarize the *steps* to customize a novel Recommender System.

Step 1. Introduce a model (with parameters and hyperparameters) to formulate the rating. For example,

$$\widehat{r}_{ui} = \mu + a_u + b_i + \mathbf{p}_u^\intercal \mathbf{q}_i$$

Step 2. Find an *algorithm* to fit the model with the observed rating set Ω and use cross-validation to find the best model:

$$\min_{\theta} \frac{1}{|\Omega|} \sum_{(u,i)\in\Omega} (r_{ui} - \widehat{r}_{ui})^2 + \lambda \mathrm{Reg}(\theta)$$

Step 3. Make a prediction based on the *best-tuned model*

» More on Matrix Factorization

MF Proposed by **Simon Funk** (2006)

Other names SVD, Matrix Factorization, Latent Factor Models

$$\hat{r}_{ui} = \boldsymbol{p}_u^{\intercal} \boldsymbol{q}_i \to r_{ui},$$

where $\mathbf{p}_u = (p_{u1}, \cdots, p_{uK})$, and $\mathbf{q}_i = (q_{i1}, \cdots, q_{iK})$.

Issue 1 No explicit regularization over users/items

Issue 2 Difficult to estimate a good latent factor for users/items with few ratings

* Can we propose a new model to address these issues?

Remark The baseline terms of all models we are going to discuss below are omitted, as they can simply be *added* or considered by *sequential fitting*.

» More MF: EDA

MovieLens:

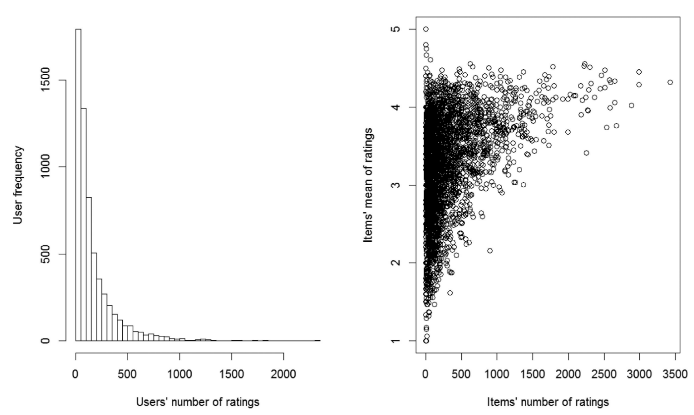


Image source: Bi, X., Qu, A., Wang, J., & Shen, X. (2017). A group-specific recommender system. Journal of the American Statistical Association, 112(519), 1344–1353.

» More MF: EDA

LastFM:

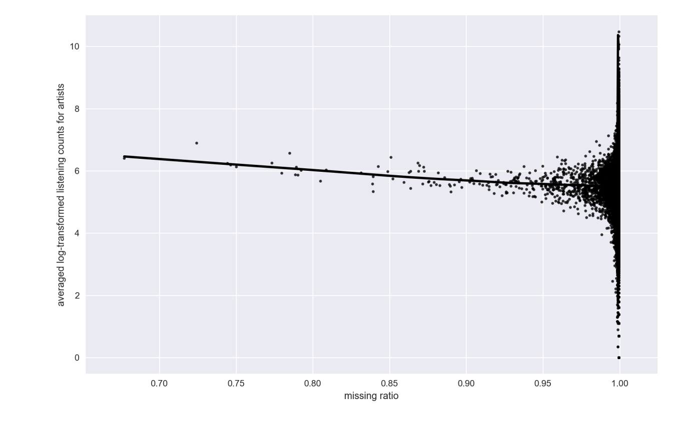


Image source: Dai, B., Wang, J., Shen, X., & Qu, A. (2019). Smooth neighborhood recommender systems. Journal of machine learning research, 20.

» More on Matrix Factorization: Exploratory Data Analysis

What conclusions can be drawn from the exploratory data analysis?

* The rating (r_{ui}) is positively correlated with *popularity*.

Popularity

  * $|\mathcal{I}_u|$ represents the number of ratings by user-u
  * $|\mathcal{U}_i|$ represents the number of ratings by item-i

Can we incorporate the popularity metrics $|\mathcal{I}_u|$ or $|\mathcal{U}_i|$, or more generally the rating patterns $\mathcal{I}_u$ or $\mathcal{U}_i$, into the **Matrix Factorization** model?

» Linear User SVD (L-SVD)

N-SVD Proposed by **Arkadiusz Paterek** (2007)

Motivation The motivation behind N-SVD is to formulate $\mathbf{p}_u$ based on $\mathcal{I}_u$ (*why?*)

- Assumes a regression model:

$$p_{uk} \sim |\mathcal{I}_u|^{1/2}|\mathcal{I}_u|^{-1}\left(w_{1k}\mathbb{I}(1 \in \mathcal{I}_u) + \cdots + w_{mk}\mathbb{I}(m \in \mathcal{I}_u)\right)$$

- (Each rated item influences the user's latent factors)
- Then $\mathbf{p}_u$ can be replaced as:

$$p_{uk} = \tau_u \sum_{i \in \mathcal{I}_u} w_{ik}, \quad \mathbf{p}_u = \tau_u \sum_{i \in \mathcal{I}_u} \mathbf{w}_i,$$

where $\tau_u = |\mathcal{I}_u|^{-1/2}$ for notational simplicity.

- The N-SVD model is:

$$\widehat{r}_{ui} = \mathbf{q}_i^\top \left(\tau_u \sum_{i' \in \mathcal{I}_u} \mathbf{w}_{i'}\right).$$

» Linear User SVD (L-SVD)

Params
- $\mathbf{Q} = (\mathbf{q}_1, \cdots, \mathbf{q}_m)^\intercal$ with $\mathbf{q}_i = (q_{i1}, \cdots, q_{iK})^\intercal$
- $\mathbf{W} = (\mathbf{w}_1, \cdots, \mathbf{w}_m)^\intercal$ with $\mathbf{w}_i = (w_{i1}, \cdots, w_{iK})^\intercal$

F Empirical loss function with regularization term:

$$\min_{\mathbf{Q},\mathbf{W}} \frac{1}{|\Omega|} \sum_{(u,i)\in\Omega} \Big(r_{ui} - \mathbf{q}_i^\intercal \big(\tau_u \sum_{i'\in\mathcal{I}_u} \mathbf{w}_{i'}\big)\Big)^2 + \lambda \sum_{i=1}^{m} \left(\|\mathbf{q}_i\|_2^2 + \|\mathbf{w}_i\|_2^2\right)$$

HP
- λ: weight of regularization term
- K: number of latent factors

» L-SVD: Discussion

Advantages

- S1 Explicitly considers the smoothness across users
- S2 Incorporates popularity into a MF model
- S3 Leverages ratings from other users to estimate latent factors for users with sparse ratings

Disadvantages

- W Less flexible in modeling users' latent factors.
- * For example, when two users have rated the same set of items, say (1,2,3), their latent factors will be identical!

Potential Solution

- Sol We can combine the strengths of SVD and L-SVD: by retaining both user-specific latent factors and item-aggregated latent factors $\rightarrow$ **SVD++**

Note that there is a trade-off between model flexibility and estimation complexity.

» SVD++

SVD++ Proposed by **Yehuda Koren** (2008)

Model SVD++ integrates SVD and L-SVD:

$$\widehat{r}_{ui} = \mathbf{q}_i^\top \Big(\mathbf{p}_u + \tau_u \sum_{j \in \mathcal{I}_u} \mathbf{w}_j\Big).$$

F Empirical loss function with regularization term:

$$\min_{\mathbf{P},\mathbf{Q},\mathbf{W}} \frac{1}{|\Omega|} \sum_{(u,i)\in\Omega} \Big(r_{ui} - \mathbf{q}_i^\top \big(\mathbf{p}_u + \tau_u \sum_{j\in\mathcal{I}_u} \mathbf{w}_j\big)\Big)^2 + \lambda \Big(\sum_{i=1}^{m} \big(\|\mathbf{q}_i\|_2^2 + \|\mathbf{w}_i\|_2^2\big) + \sum_{u=1}^{n} \|\mathbf{p}_u\|_2^2\Big)$$

Params

- User latent factors: **P**
- Regression latent factors: **W**

HP

- λ: regularization strength
- K: number of latent factors

» Nonnegative Matrix Factorization (NMF) (Optional)

NMF Introduced by **Paatero** and **Tapper** (1994), and further developed by **Lee** and **Seung** (1999)

Model The predicted rating is given as:

$$\hat{r}_{ui} = \boldsymbol{p}_u^\top \boldsymbol{q}_i \approx r_{ui}$$

Formulation Empirical loss function with nonnegative constraints:

$$\min_{\mathbf{P},\mathbf{Q}} \frac{1}{|\Omega|} \sum_{(u,i)\in\Omega} (r_{ui} - \boldsymbol{p}_u^\top \boldsymbol{q}_i)^2 + \lambda \mathrm{Reg}(\mathbf{P},\mathbf{Q}), \quad \mathbf{P} \geq \mathbf{0}; \mathbf{Q} \geq \mathbf{0}$$

» Group SVD (gSVD)

Group MF Introduced by **Bi, et al.** (2017)

Grouping Pre-computed group assignments for users and items

- User Group Each user u is assigned to a group with ID ν_u
- Item Group Each item i is assigned to a group with ID j_i

Model The predicted rating is given as:

$$\hat{r}_{ui} = (\boldsymbol{p}_u + \mathbf{s}_{\nu_u})^\intercal (\boldsymbol{q}_i + \mathbf{t}_{j_i}) \approx r_{ui}$$

» Smooth SVD (sSVD)

sSVD Introduced by **Dai, et al.** (2019)

Weight Pre-computed similarity weights between pairs of users and items

- User Sim $w_{u,u'}$ represents the similarity between users u and u'
- Item Sim $w_{i,i'}$ represents the similarity between items i and i'

Model The predicted rating is given as: $\hat{r}_{ui} = \mathbf{p}_u^\intercal \mathbf{q}_i \approx r_{ui}$

F Using *neighborhood* ratings to estimate (u, i), with weighted regularization:

$$\min_{\mathbf{P},\mathbf{Q}} \frac{1}{nm} \sum_{u=1}^{n} \sum_{i=1}^{m} \sum_{(u',i') \in \Omega} w_{u,u'} w_{i,i'} (r_{u'i'} - \mathbf{p}_u^\intercal \mathbf{q}_i)^2 + \lambda \mathrm{Reg}(\mathbf{P}, \mathbf{Q})$$

Advantages of SVD-/MF-based Methods

+ Widely adopted in both industry and academic research
+ Exhibits competitive performance
+ Amenable to efficient implementation using parallel computing architectures
+ Achieves *state-of-the-art (SOTA) performance* when only rating data is available
- Most models suffer from *cold-start* users and/or items problems