

# Review of Collaborative Filtering in Recommender System

Final Project Report of CS 6241

Wentao Guo

## Abstract

## Keyword

SVD, SVD++, Recommender System, Collaborative Filtering

## 1 Background

Developed in 1990s, the recommender system (RS) has been applied in e-commerce, music apps, job portal, social networking and more [2]. The recommender system collects information on users' behaviors explicitly (users' ratings on items) and implicitly (users' mouse movements, attention on one page, etc.) and makes prediction on users' preferences. For example, Netflix applies a five-star ratings to help users find their favorite movies and maintain their subscriptions [4], and Amazon selects products based on the predicted users' favors to gain profits.

In October 2006, Netflix launched a competition for which they rewarded the team that can beat Netflix's Cinematch system by at least 10% Root Mean Squared Error [4]. This competition arouse great attention in collaborative filtering field as the dataset covered 100 million ratings for 500, 000 anonymous customers on 17, 000 movies, which was greater than previous public dataset in the orders of magnitude [11]. The final grand prize was given to "Bellkor's Pragmatic Chaos" team in 2009 [4] [10].

In this competition, matrix factorization approaches

raised people's attention as the top teams in this competition frequently applying it to beat Bayes or probabilistic approaches [9]. Singular Value Decomposition as the classical latent semantic indexing approach in information retrieval was found as a great fit for spanning customers data with low dimensions and often adapted to various variants as regularized SVD, SVD++, iterative SVD, and more [4] [7] [6] [9].

## 2 Rationale behind Collaborative Filtering Recommender System

### 2.1 User-item Interaction

Matrix factorization models represents the user-item interaction in a latent joint space [11]. Each item  $i$  is associated with a vector  $q_i \in \mathbb{R}^f$  and user  $u$  is associated with a vector  $p_u \in \mathbb{R}^f$ . The rationale behind such representation is: item  $i$  will possess some unknown factors in a measure, positive or negative, stored in entries of  $q_i$ , and user will have different favors toward these factors, and such interest is represented as entries in  $p_u$ . Therefore, we can take inner product to get an approximate of the user  $u$ 's rating on item

$i$  as [11]

$$\hat{r}_{ui} = q_i^T p_u \quad (1)$$

Conventional SVD suffers from an incomplete (or even sparse) rating matrix. Researchers often resort to imputation to fill in the matrix, or through regularized least squared model as [11].

$$\min_{q,p} \sum_{(u,i) \in \kappa} (r_{ui} - q_i^T p_u)^2 + \lambda(\|q_i\|^2 + \|p_u\|^2) \quad (2)$$

in which  $\kappa$  is the set of known ratings. The regularizing term  $\|q_i\|^2 + \|p_u\|^2$  controls the magnitude of  $q_i$  and  $p_u$  to avoid overfitting.

## 2.2 Bias

Early recommender system exhibit a systematic bias for users giving generally higher ratings than others or items receive higher ratings than others [11]. Therefore, as Koren, et al. suggested, the entire user-item interaction cannot be simply explained by  $q_i^T p_u$ , and a bias term exists. A first-order approximation of bias is as follows [11]

$$b_{ui} = \mu + b_i + b_u \quad (3)$$

in which  $b_{ui}$  is the bias and  $\mu$  is the overall average rating of training data, and  $b_u, b_i$  are user  $u$  and item  $i$ 's average ratings compared to the overall average rating [11]. We can transform our estimated rating as

$$\hat{r}_{ui} = b_{ui} + q_i^T p_u \quad (4)$$

and the objective becomes

$$\min_{q,p} \sum_{(u,i) \in \kappa} (r_{ui} - q_i^T p_u - b_{ui})^2 + \lambda(\|q_i\|^2 + \|p_u\|^2 + b_u^2 + b_i^2) \quad (5)$$

## 3 SVD-Related Methods

### 3.1 Standard SVD-CF Model

SVD, as a common dimension reduction tool, originally used in information retrieval to identify latent semantic factor (as Latent Semantic Indexing), can represent the user-item

interactions in a low dimensional space [11] [15]. SVD approximates a matrix  $A \in \mathbb{R}^{m \times n}$  in a form as

$$A = U \Sigma V^T \quad (6)$$

Suppose the rank of  $A$  is  $r$ , there are three main properties in SVD[15] [4]:

- The initial  $r$  singular values of diagonal matrix  $\Sigma$  holds that  $\forall i \in [1, r] \sigma_i > 0$  and  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ .
- The first  $r$  columns of orthogonal matrix  $U$  are eigenvectors of  $A A^T$  and span the column space of  $A$ .
- The first  $r$  columns of orthogonal matrix  $V$  are eigenvectors of  $A^T A$  and span the row space of  $A$ .

People usually takes the best rank- $k$  approximation of matrix  $A$  as

$$A_k = U_k \Sigma_k V_k^T \quad (7)$$

with  $U_k \in \mathbb{R}^{m \times k}$ ,  $\Sigma_k \in \mathbb{R}^{k \times k}$ ,  $V_k \in \mathbb{R}^{k \times n}$ .

We will take a rating matrix  $R$  with users  $u_1, u_2, \dots, u_i$  in the row and items  $i_1, i_2, \dots, i_j$  in the column, and the entry  $R_{ij}$  is the estimated interest of user  $u_i$  on item  $i_j$ . [11] [8] [15].

Notice that matrix  $R$  is often sparse and require some preprocessing. A common preprocessing approach is to take the row averages to replace all of the missing values in the matrix  $R$  and normalize  $R$  by subtracting row averages [15] [8].

We then compute  $U_k \Sigma_k V_k^T$  as equation 7, and form two matrix products:  $U_k \sqrt{\Sigma_k^T}$  and  $\sqrt{\Sigma_k} V_k^T$ .

There are some variants regarding the computation of predicted rating [15] [14]. For example, Manolis et al. [14] provided a formula for SVD-CF prediction rating for user  $u_i$  on item  $i_j$  as:

$$P_{ij} = \bar{r}_i + U_k \sqrt{\Sigma_k^T[i]} \sqrt{\Sigma_k} V_k^T(j) \quad (8)$$

Another variant is based on PCA, with the previous steps identical except that we need to replace and subtract column average  $\bar{c}_j$  for preprocessing  $R$ . In addition, prediction score for user  $u_i$  on item  $i_j$  will be

$$P_{ij} = \bar{c}_j + U_k \sqrt{\Sigma_k^T[i]} \sqrt{\Sigma_k} V_k^T(j) \quad (9)$$

Some researchers incorporate the reduced matrix  $R_{\text{red}} = U_k \Sigma_k V_k^T$  into the calculation of similarity measure to predict the new user or item's score based on the existing clusters. These approaches are usually categorized to user-based collaborative filtering or item-based collaborative filtering depending on the subject of similarity measure.

Pearson correlation and adjusted cosine similarity are common similarity measures in literature [13][15] [12] [1] [3]:

- Pearson correlation

$\mathcal{I}_{uv}$  means common item rated by both user  $u$  and user  $v$ .  $\bar{r}_u$  means average rating of user  $u$  on the shared item set.

$$\text{sim} = \frac{\sum_{i \in \mathcal{I}_{uv}} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in \mathcal{I}_{uv}} (r_{ui} - \bar{r}_u)^2 (r_{vi} - \bar{r}_v)^2}} \quad (10)$$

- Adjusted cosine similarity

$\bar{r}_u$  means average rating of user  $u$  on the entire item set.

$$\text{sim} = \frac{\sum_i (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_i (r_{ui} - \bar{r}_u)^2 (r_{vi} - \bar{r}_v)^2}} \quad (11)$$

Pearson-correlation looks really similar to adjusted cosine similarity except that adjusted cosine similarity is calculated over the entire set of rated vectors while person correlation is calculated over co-rated vector [5]. For the missing value case, a typical approach is to set it as zero [12].

After we calculate similarity measure, we will

### 3.2 SVD++ Model

The traditional SVD method only considers users' explicit ratings but do not take implicit data into account. Although implicit data (mouse move, reading time, etc.) might not be publicly available, there is one kind of implicit data always available: whether a user rates on an item (rate vs. not rate).

SVD++ model was developed by Koren [9]. The complete formula is shown below

$$\begin{aligned} \hat{r}_{ui} = & \mu + b_i + b_u \\ & + q_i^T (p_u + |N_u|^{-.5} \sum_{j \in N_u} y_j) \\ & + |R^k(i; u)|^{-.5} \sum_{j \in R^k(i; u)} (r_{uj} - b_{u_j}) w_{ij} \\ & + |N^k(i; u)|^{-.5} \sum_{j \in N^k(i; u)} c_{ij} \end{aligned} \quad (12)$$

Rule 12 is a 3-tier model. The first tier is  $b_{ui}$  as mentioned before.

The second tier  $q_i^T (p_u + |N_u|^{-\frac{1}{2}} \sum_{j \in N(u)} y_j)$  describes the intersection between user profile and item profile. We need to introduce a few notations to understand this tier:

- $R_u$  all items for which user  $u$  provides a rating
- $N_u$  all items for which user  $u$  provides an implicit preference
- $S^k(i)$   $k$  items rated by  $u$  that are similar to item  $i$
- $N^k(i; u) = N_u \cap S^k(i)$
- $R^k(i; u) = R_u \cap S^k(i)$
- $w_{ij}$  weight from  $j$  to  $i$
- $c_{ij}$  significance offset (indication of predictability by  $j$  to  $i$ )

The remaining terms form the final tier, or "neighborhood tier" called by Koren, as it contributes some adjustments to the profile.

### 3.3 Regularized SVD Model

## 4 Experiment

similarity

## 5 Conclusion

## References

- [1] Lamis Al Hassanieh, Chadi Abou Jaoudeh, Jacques Bou Abdo, and Jacques Demerjian. Similarity measures for

- collaborative filtering recommender systems. In *2018 IEEE Middle East and North Africa Communications Conference (MENACOMM)*, pages 1–5. IEEE, 2018.
- [2] James Bennett, Stan Lanning, et al. The netflix prize. In *Proceedings of KDD cup and workshop*, volume 2007, page 35. Citeseer, 2007.
- [3] Jesús Bobadilla, Antonio Hernando, Fernando Ortega, and Abraham Gutiérrez. Collaborative filtering based on significances. *Information Sciences*, 185(1):1–17, 2012.
- [4] Stephen Gower. Netflix prize and svd, 2014.
- [5] Lin Guo and Qin Ke Peng. A combinative similarity computing measure for collaborative filtering. In *Applied Mechanics and Materials*, volume 347, pages 2919–2925. Trans Tech Publ, 2013.
- [6] Maryam Jallouli, Sonia Lajmi, and Ikram Amous. When contextual information meets recommender systems: extended svd++ models. *International Journal of Computers and Applications*, pages 1–8, 2020.
- [7] Yancheng Jia, Changhua Zhang, Qinghua Lu, and Peng Wang. Users’ brands preference based on svd++ in recommender systems. In *2014 ieee workshop on advanced research and technology in industry applications (wartia)*, pages 1175–1178. IEEE, 2014.
- [8] Dohyun Kim and Bong-Jin Yum. Collaborative filtering based on iterative principal component analysis. *Expert Systems with Applications*, 28(4):823–830, 2005.
- [9] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434, 2008.
- [10] Yehuda Koren. The bellkor solution to the netflix grand prize. *Netflix prize documentation*, 81(2009):1–10, 2009.
- [11] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.
- [12] Haifeng Liu, Zheng Hu, Ahmad Mian, Hui Tian, and Xuzhen Zhu. A new user similarity model to improve the accuracy of collaborative filtering. *Knowledge-Based Systems*, 56:156–166, 2014.
- [13] Ritu Sharma, Dinesh Gopalani, and Yogesh Meena. Collaborative filtering-based recommender system: Approaches and research challenges. In *2017 3rd international conference on computational intelligence & communication technology (cict)*, pages 1–6. IEEE, 2017.
- [14] Manolis Vozalis, Angelos Markos, and Konstantinos Margaritis. On the performance of svd-based algorithms for collaborative filtering. In *2009 Fourth Balkan Conference in Informatics*, pages 245–250. IEEE, 2009.
- [15] Manolis G Vozalis and Konstantinos G Margaritis. Applying svd on generalized item-based filtering.