Learning from Sets

***Abstract*—Existing item recommendation methods rely on the availability of the user's preferences on the items consumed in the past. Additionally, the performance of these recommendation methods is proportional to the amount of information available to the recommender system, i.e., the more availability of user's preferences on the items, the better the quality of the recommendations for the user. However, it is not easy to elicit the user's preferences on items in order to improve the recommendations for the user, particularly when we have a large number of items. In this paper, we propose a method in which we elicit the user's preference on a *set* of items rather than on the individual items that constitute the *set*. We use this preference on the set to estimate the user's preference on the items that constitute the set. This method, compared to the process in which the user has to sequentially provide his/her preference for the individual items, can quickly gather the user's preferences on a few sets that span a large number of items. In the experiments on real world data, we show that the quality of recommendations generated from these sets is comparable to the recommendations generated using the individual preference on the items.**

I. Introduction

Recommender systems are widely deployed in various domains e.g. music (iTunes), movies (Netflix) and e-commerce (Amazon) to suggest relevant products to the users based on their preferences.

One of the key methods used by the recommender systems is the collaborative filtering method, which uses the information from the previous transactions of the users and suggest relevant products to consume in the future. Generally, the collaborative filtering consists of two approaches: the neighborhood - based approach and the latent-factor models. In the neighborhood - based approach [2], the similarities between the items or the users are used to compute recommendations. While the latent-factor models [1] transform the users and the items to the same latent space where both of them are comparable to each other, the items closest to the users in this space serve as the recommendations to the users.

An important requirement of these collaborative filtering methods is that the user should indicate his/her preference over a certain number of items to generate recommendations for future consumption. For existing users this is not a significant problem, as these users have already consumed some items in the past and will continue to provide information as they consume the items. However, this is a significant problem for the new user as there is no prior information available about their preferences and hence the recommender systems fails to provide personalized recommendations to such users. These users are also known as "cold-start" users.

To overcome this problem for a new user, most services deploying recommender systems elicit a user's preferences on a few popular items which the user may have consumed before. This process of eliciting the preferences of the users on the items is time consuming as a user has to indicate his/her preferences on items one at a time. Hence, the user may quit this preference elicitation process without completing it, thus leading to poor initial recommendation performance for the user.

In this work, we present a method which elicits preferences of the user on a set of items rather than rating those items individually in a sequence. This preference on the set is used to estimate the preferences on the individual items that constitute the set. Since a set can be composed of multiple items the user can indicate his/her preference on a large number of items by indicating his/her preference on a few sets of items. Hence, the time spent by a cold-start user to elicit his preferences in order to get personalized recommendations is reduced significantly.

We make the following contributions: First, we propose various Learning from Set models that enables estimation of the user's preference on the individual items. Second, we illustrate the effectiveness of these models with experiments on the synthetic and the real datasets by comparing their performance with the state-of-the-art matrix factorization methods.

The rest of the paper is organized as follows. Section 2 introduces the notations used in the paper. In Section 3, the relevant existing methods are described. Section 4 presents the Learning from Set models. In Section 5, details about evaluation methodology and dataset are provided. Sections 6 provides the results of the experimental evaluation. Finally, Section 7 gives some concluding remarks.

II. Notations and Definitions

Throughout the paper, all vectors are column vectors and are represented by bold lowercase letters (e.g., $\boldsymbol{u}$). Matrices are represented by upper case letters (e.g., R, U, V).

The historical preference information is represented by a preference matrix R. Each row in R corresponds to a user and each column corresponds to an item. The entries of R indicates the users' preferences on the items. The preference given by the user u for the item i is represented by entry $r_{u,i}$ in R. The symbol $\tilde{r}_{u,i}$ represents the score predicted by the model for the actual preference $r_{u,i}$.

Sets are represented with calligraphic letters. The set of items $\mathcal{S}$ has size $|S|$.

III. Related Work

A. Collaborative Filtering

Collaborative filtering is one of the widely used methods in recommender systems. It tries to estimate the rating on a user u on item i i.e., $\tilde{r}_{ui}$ based on the partially observed user-item rating matrix $R \in \mathcal{R}^{m \times n}$ for m users and n items. In order to generate recommendations of new items, we need to estimate the unobserved entries of the matrix R. The unobserved entries can be estimated by assuming the matrix R to be of low-rank and completing the matrix R by minimizing a squared loss:

$$(\tilde{P}, \tilde{Q}) = \arg\min_{P,Q} \sum_{(u,i)} (R - [PQ^T]_{u,i})^2, \quad (1)$$

where $P \in R^{m \times r}, Q \in R^{n \times r}$. The completed matrix $\tilde{R} = \tilde{P}\tilde{Q}^T$ is used to serve the recommendation to the user for the items for which his preferences were unknown in the original matrix R.

IV. Methods

To estimate the preferences on individual items from the preference of the user on the sets, we need to understand how the user rate a given set of items. A user may rate a set of items by considering each item in the set or may chose to rate the set of the items based on few but majority of the items in the set.

In the first case, when a user consider all the items in the set before assigning a rating to a set then we can safely assume that the user most likely gives the set an average score of his preferences for all the items that constitutes the set. Under this assumption the estimated rating of the user u on a set S is given by:

$$\tilde{r}_{us}^{avg} = \frac{1}{|S|} \sum_{i \in \mathcal{S}} r_{u,i}, \quad (2)$$

where $\mathcal{S}$ denotes the set containing the items and $r_{u,i}$ is the preference score of the user u on the item i .

Similarly, when a user considers only a few but the majority of the items in the set $\mathcal{S}$ then the user's estimated rating on the set $\mathcal{S}$ is given by:

$$\tilde{r}_{us}^{maj} = \frac{1}{|\mathcal{S}_{maj}|} \sum_{i \in \mathcal{S}_{maj}} r_{u,i}, \quad (3)$$

where $\mathcal{S}_maj$ contains the majority of the top rated items in the set $\mathcal{S}$.

Assuming that the original user-item preference matrix R is low-rank, we can write the estimated preference of the user u for the item i as follow:

$$\tilde{r}_{ui} = p_u^T q_i, \quad (4)$$

where, k is the rank of the matrix R, $p_u \in \mathcal{R}^k$ and $q_i \in \mathcal{R}^K$ denotes the latent factor of the user u and the item i respectively.

Following the low-rank assumption, we can rewrite the estimated score of the user u for a set $\mathcal{S}$ in equations 2 and 3 as follow:

$$\tilde{r}_{us}^{avg} = \frac{1}{|S|} \sum_{i \in \mathcal{S}} p_u^T q_i, \quad (5)$$

$$\tilde{r}_{us}^{maj} = \frac{1}{|\mathcal{S}_{maj}|} \sum_{i \in \mathcal{S}_{maj}} p_u^T q_i, \quad (6)$$

A. Learning from Set Model

The learning from set model is parameterized by $\Theta = [P, Q]$, where the matrices P and Q contains the latent factors of the users and the items respectively. The model parameters are estimated by minimizing the squared error loss function, given by

$$\mathcal{L}_{rmse}(\Theta) \equiv \sum_{u \in U} \sum_{s \in \mathcal{R}_{us}} (\tilde{r}_{us} - r_{us})^2, \quad (7)$$

where $\mathcal{R}_{us}$ contains all the sets preferred by the user u, r_{us} is the original rating and $\tilde{r}_{us}$ is the estimated rating of the user u on the set s. The estimated rating $\tilde{r}_{us}$ can be computed as per equation 5 or 6 depending on whether the average or the majority assumption is used to estimate the rating of the user u on the set s.

To control model complexity, we add regularization of the model parameters thereby leading to an optimization process of the following form:

$$\min_{P,Q} \sum_{u \in U} \sum_{s \in \mathcal{R}_{us}} (\tilde{r}_{us} - r_{us})^2 + \lambda(\|P\|_F^2 + \|Q\|_F^2), \quad (8)$$

where λ is the regularization parameter.

The optimization problem of equation 8 is solved by stochastic gradient descent algorithm for the average assumption and stochastic sub-gradient descent algorithm for the majority assumption.

V. Experimental Evaluation

VI. Results and Discussion

VII. Conclusion

References

[1] 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*, KDD '08, pages 426–434, New York, NY, USA, 2008. ACM.

[2] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web*, WWW '01, New York, NY, USA, 2001. ACM.