PROGRESS REPORT

Group Recommendation System using Matrix Factorization

By
Rajesh, Roll No. 177924 MCA-IIIrd
Project Guide: Dr. Venkateswara Rao Kagita

INTRODUCTION

Recommender systems aim to provide information items that are of potential interest to a user. There are certain scenarios in which recommending a set of items to a group of several users is more appropriate than providing several sets of recommendations to each individual user of the group. Group Recommender Systems aim to provide a set of recommendations that satisfies the preferences of all users in a group. Matrix factorization based collaborative filtering is the most popular and successful approach for personalized recommender systems. The key idea of matrix factorization (MF) based techniques is to learn low dimensional user latent factors (U) and item latent factors (V) to simultaneously approximate the observed entries under some loss measure and predict the unobserved entries. We accomplish prediction by determining U×V.

In this work, we extend the concept matrix factorization to the group recommendation scenario. We decompose the rating matrix into four matrices that represent users, items, user groups and item groups in a latent factor space. These matrices are later used to recommend items or group of items as a package to user or group of users.

RECOMMENDATION METHOD

In this section, we present matrix factorization based approach for group recommendation. The main idea of MF models is to factorize the original rating matrix into two matrices that represent user-item interactions. The key factor of the proposed method is to factorize the original matrix into four matrices representing users, items, user groups and item groups' latent factors. MF models map users and items to a joint latent factor space to represent user-item interactions. The model used in this paper factorizes the rating matrix into the following elements:

- $U_i = (U_{i,1} .....U_{i,d})$ represents the factor vector of the user i .
- $V_j = (V_{j,1} .....V_{j,d})$ represents the factor vector of the item j .
- $U^G{}_u = (U^G{}_{u,1} .....U^G{}_{u,d})$ represents the factor vector of the user group G_u .
- $V^G{}_g = (V^G{}_{g,1} .....V^G{}_{g,d})$ represents the factor vector of the item group G_g .

To learn the factor vectors U_i , V_j , U^G_u and V^G_g the system minimizes the following expression for a set of known ratings:

$$\min_{U,V,U^G,V^G} J(U,V,U^G,V^G) = \sum_{i,j\in\Omega} (Y_{i,j} - U_i V_j) + \lambda_1 \sum_{i=1}^{M} \left\| U^G_{I_i} - U_i \right\|^2 + \lambda_2 \sum_{j=1}^{N} \left\| V^G_{J_j} - V_j \right\|^2 + \lambda_3 \|U\|^2 + \lambda_4 \|V\|^2$$

Where $Y_{M\times N}$ is a user-item rating matrix wherein *M* is the number of users and *N* is the number of items. *K1* and K2 are the number of user and item groups respectively. Ω is the set of observed entries, I_i represents the group of a user *i,* and J_j represents the group of an item *j*. λ_1, λ_2, λ_3 and λ_4 are parameters that control the training process. The matrices are initialized randomly. To minimize the error, we have to know in which direction we have to modify the values of U_{ik}, V_{kj},U^G_{uk} and V^G_{gk}. In other words, we need to know the gradient at the current values. Therefore, we differentiate the above equation with respect to U_{ik}, V_{kj},U^G_{uk} and V^G_{gk} separately. After obtaining the gradient, we update U_{ik}, V_{kj},U^G_{uk} and V^G_{gk} as follows.

$$\mathbf{U_{ik}} = \mathbf{U_{ik}} + \alpha((\mathbf{Y_{i,j}} - \mathbf{U_{ik}} \mathbf{V_{kj}})\mathbf{V_{kj}} + \lambda_1 (\mathbf{U^G_{I_i k}} - \mathbf{U_{ik}}) - \lambda_3 \mathbf{U_{ik}})$$

$$\mathbf{V_{kj}} = \mathbf{V_{kj}} + \alpha((\mathbf{Y_{i,j}} - \mathbf{U_{ik}} \mathbf{V_{kj}})\mathbf{U_{ik}} + \lambda_2 (\mathbf{V^G_{kJ_j}} - \mathbf{V_{kj}}) - \lambda_4 \mathbf{V_{kj}})$$

$$\mathbf{U^G_{uk}} = \mathbf{U^G_{uk}} - \alpha\lambda_1 \sum_{\substack{\mathbf{u=I_i} \\ \forall\, \mathbf{i\in(0\,,M)}}} (\mathbf{U^G_{uk}} - \mathbf{U_{ik}})$$

$$\mathbf{V^G_{gk}} = \mathbf{V^G_{gk}} - \alpha\lambda_2 \sum_{\substack{\mathbf{g=J_j} \\ \forall\, \mathbf{j\in(0\,,N)}}} (\mathbf{V^G_{gk}} - \mathbf{V_{jk}})$$

Where α is a parameter that controls the speed of the learning process.

EXPERIMENTAL ANALYSIS

In order to validate the group recommender system we have used the MovieLens 1M dataset. We experiment our model for groups sizes in the range of [2, 12]. These groups were generated randomly. We select 30% of ratings as test set and rest are used for training. In each experiment, we generate a set of recommendations for each group of users and computed their precision and recall. We compare the precision for top 20 recommendations with existing recommendation models. In the comparison between the traditional KNN based collaborative filtering and the proposed MF based collaborative filtering for groups of users, there is a significant increase in the quality of recommendations. For smaller groups(2-4), we got average precision 0.67 and for mid-size (6-8) and large size groups (10-12), we got average precision 0.69.

The expected results of this model is to improve the result in existing MF based collaborative filtering which computes the group factors by merging the factors of the users that belong to the group.

PROGRESS

- Implemented a Matrix Factorisation model to generate users, items, user groups, item groups latent factor for Group Recommendation.
- Generating the Recommendations Items to user, Items to user group.
- Mean Absolute Error, Root Mean Square Error, Precision and Recall are used to evaluate the recommendation quality.
- Comparing the results by generating random groups and generating groups using clustering techniques.
- Estimating the MF parameters by comparing the results and selecting the one providing the best result.
- Experiment Setup consists of MovieLens 1M dataset in which I had compared the Precision and Recall for different no of group sizes, no of factors and no of recommendations.

ADDITIONAL WORK

Some tasks that are to be completed in the coming week:

- Recommendation of Item groups to user and user groups.
- Implementation of State of Art methods and comparative study.

In order to validate the group recommender system we have used the MovieLens 1M dataset. The MovieLens database contains 1,000,209 anonymous ratings of 3,706 movies made by 6,040 users who joined MovieLens in 2000. We estimate the latent factor of the MF model to use a learned model for recommendation predictions. We use the data to estimate the MF parameters and select the one providing the best results. Table 8 shows the parameters used to factorize dataset. MovieLens datasets doesn't contain information about how the users and items are grouped, so we generated random groups of users and items fixing their sizes from 2 to 12. . We selected 30% of ratings as test ratings and these ratings were not used during the learning phase. In each experiment we generated the set of recommendations for each group of users and we computed their precision and recall. We averaged the evaluation metrics values based on the group size. We set three different types of groups: small groups (2 to 4), mid-size groups (5 to 8) and large groups (9 to 12). We aim to analyze the impact of the group size on the quality of recommendations. To avoid fluctuations generated by the random selection of the groups and the random initialization of the latent factors of the MF model we repeated each experiment multiple times and averaged the results.