



Explainable Matrix Factorization with Constraints on Neighborhood in the Latent Space

Shuo Wang, Hui Tian^(✉), Xuzhen Zhu, and Zhipeng Wu

State Key Laboratory of Networking and Switching Technology,
Beijing University of Posts and Telecommunications, Beijing 100876, China
{wangshuo16,tianhui,zhuxuzhen,wzp}@bupt.edu.cn

Abstract. Nowadays, recommender systems are widely used to solve the problem of information overload in modern society. And most of the previous studies focus overwhelmingly on high accuracy in the recommender systems. But in a real system, the high accuracy does not always satisfy overall user experience. The explainability has a great impact on the user experience. We mainly focus on the explainability of recommender systems in this paper. To the best of our knowledge, it is the first time that the neighborhood information in the latent space is integrated into the Explainable Matrix Factorization. We change the method of calculation of the explainability matrix and consider the neighbors' weight to further improve performance. We use the benchmark data set (MovieLens) to demonstrate the effectiveness of the proposed Neighborhood-based Explainable Matrix Factorization. And the result shows a great improvement for accuracy and explainability.

Keywords: Recommender systems · Explainability
Matrix factorization · Neighborhood information

1 Introduction

Information overload has become a serious obstacle with the development of Internet technology. Recommender systems provide a promising way to alleviate the dilemma. Recommender systems become more accurate with the deepening of machine learning research applied in it. However, prediction accuracy is not the only metric that should be taken into consideration in recommender systems. Many other metrics are equally important. One of them is the explainability which is used to describe the reason why the items are recommended to users and show the transparency of recommender systems. The explanation of recommender systems plays important roles in the industries, e.g., helping users make better choices, enhancing customers' trust to e-commerce web sites. Therefore, what we recommend and how we recommend are equally important to improve the performances of recommender systems [3].

There are many approaches to show explanation in the recommender systems. For example, there are some labels such as “customers who bought this item also bought that...” that are shown to customer on the TaoBao. In the article [6], Gedikli et al. made use of semi-structured interviews in which authors asked the participants about their opinions towards explanation. In the article [14], Jesse et al. proposed community tag to achieve explanation which was mainly based on content model. Zhang et al. [15] proposed an explicit factor model based on Phrase-level Sentiment Analysis. Abdollahi et al. [2] proposed a method (EMF), which only used the rating data and incorporate the neighborhood-based CF method and latent factor models. Explainability could be directly formulated based on the rating distribution within the user’s or item’s neighbors, providing a basis upon which to explain the recommendations.

The EMF only uses a simple expected value of the ratings of items to get the explainability which is not very reasonable. Besides it still directly factorizes user-item interactions in the latent space, which is poor at identifying strong associations among a small set of closely related items [9], especially when data is highly sparse. Based on those considerations, we change the method of calculation of explainability and enhance EMF’s ability to leverage localized information by joining the neighborhood information in the latent space.

Our contributions are summarized as follows: (1) We change the method of calculation to get the explainability matrix more reasonable; (2) It is the first time to integrate the neighborhood information in the latent space into the explainable matrix factorization; (3) We add the neighbors’ weight in the explainability constraint term; (4) Experiments on real-world data set demonstrate the effectiveness of our model for the explainable recommendation task in the situation that the absence of any additional data source, such as item content or user attributes.

2 Preliminary

2.1 Explainability

As we all known, the recommendation list obtained by the neighborhood-based CF method is explainable. We can explain the recommended result to customers like “your friends bought them, and they gave a high comprehensive rating, so we recommend them to you.” So we can make use of it to achieve explanation. For the neighborhood-based explanation, it is divided into user-based explanation and item-based explanation which are analogous. Take user-based explanation for example. Given the set of similar users for user u , the empirical density distribution of the similar users ratings on the recommended item i can be obtained by dividing the counts for each rating value by the total counts. It equals to the empirical conditional probability of ratings of item i [2]. For each rating m in the set of ratings, the probability is written as follows:

$$Pr(r_{v,i} = m | v \in N_u) = \frac{|N_u \cap U_{i,m}|}{N_u}, \quad (1)$$

where $r_{v,i}$ denotes the rating of user v to item i ; N_u is the set of similar users of user u and $U_{i,m}$ stands for the set of users who have given rating m to item i . Through Eq. 1, we can get the expected value of rating given by the similar users to the recommended item i . A reasonable and intuitive measure of explanation can be given by the expected value as follows:

$$E(r_{v,i}|N_u) = \sum_{m \in M} m * P_r(r_{v,i} = m|v \in N_u). \quad (2)$$

Item-based formulas have the similar expression with Eqs. 1 and 2.

In this study, considering the difference of neighbors, the expected value can be obtained more reasonable through similarity, such as Pearson Correlation Coefficient of users (items) which is more reasonable. The equation is given as follows:

$$Sim(u, v) = \frac{\sum_{i \in I_{uv}} (r_{ui} - \overline{R_u}) * (r_{vi} - \overline{R_v})}{\sqrt{\sum_{i \in I_{uv}} (r_{ui} - \overline{R_u})^2 * \sum_{i \in I_{uv}} (r_{vi} - \overline{R_v})^2}}, \quad (3)$$

where I_u is the set of items rated by user u , I_v is the set of items rated by user v and $I_{uv} = I_u \cap I_v$; $\overline{R_u}$ and $\overline{R_v}$ is the average of rating of user u and user v , respectively.

$$E(r_{v,i}|N_u)_{sim} = \frac{\sum_{v \in N_u} r_{v,i} * Sim(u, v)}{\sum_{v \in N_u} |Sim(u, v)|}. \quad (4)$$

Because of the different relationships between every user (item), it is inaccurate that we only use the empirical conditional probability of rating to get the expected rating result. Considering each individual's effect on this active user (item) is more reasonable. So in this study, we use the $E(r_{v,i}|N_u)_{sim}$ as a soft constraint in cost function that can achieve a mixture of explanation and recommendation.

2.2 Explainability Matrix

To represent the relationship of users and items, a bipartite graph G is used. The edge weights can generate the explainability matrix W whose values represent the explanation of items to users. If the value of W for item is larger, the item is more easily explained. Otherwise, the case is opposite. Hence the explainability matrix W is defined as follows:

$$W_{u,i} = \begin{cases} E_{u,i} & \text{if } E_{u,i} \geq \theta \\ 0 & \text{otherwise} \end{cases}, \quad (5)$$

where $E_{u,i}$ is either $E(r_{v,i}|N_u)$ or $E(r_{v,i}|N_i)$, depending on the specific case (based on users or based on items), and θ denotes a threshold set above which we can accept that item i is explainable for user u . θ is set by ourself, depending on our consciousness.

2.3 Explainable Matrix Factorization

Matrix Factorization transforms original rating matrix $R_{n \times M}$ into two lower-rank matrices P and Q . The rank of P and Q is defined as f . The rating is predicted by the approximation of $R_{n \times M} \approx P_{n \times f} Q_{f \times m}^T$. User u and item i are associated with the factors $p_u \in \mathbb{R}^f$ and $q_i \in \mathbb{R}^f$, respectively. Each item and Each user are mapped to factor vectors $p_u, q_i \in \mathbb{R}^f$ by making use of machine learning method, respectively. The explainable MF [2] is obtained by adding explainability as a constraint to matrix factorization:

$$J = \sum_{u,i \in R} (r_{u,i} - p_u q_i^T)^2 + \frac{\beta}{2} (\|p_u\|^2 + \|q_i\|^2) + \frac{\gamma}{2} \|p_u - q_i\|^2 W_{u,i}. \quad (6)$$

For the last term (explainability constraint term) of the formula, we can illustrate it as follow:

For the explainability parameter $W_{u,i}$, when $E_{u,i}$ is lower than a certain threshold, we think that the item is unexplainable, so the $W_{u,i}$ equals zero, and the last term is zero. The formula becomes the loss function of the MF. When $E_{u,i}$ exceeds a certain threshold, we think that the item is explainable, and the $W_{u,i}$ equals the $E_{u,i}$ (same as the prediction rating based on CF). The greater the value is, the stronger the explainability is.

The explainability has some influence on the latent space. The p_u represents the features of user u and the q_i represents the features of item i in the latent space. Only when those features are similar (close in the space), the predicted rating of item i for user u will be high. So, if the recommended item is explainable, which means that the predicted rating of item is high (high rating item should be recommended), the p_u and q_i will be close in the latent space. Thus, when we minimize the objective function J , we can ensure items that have higher explainability relation ($W_{u,i}$ is large) to a user, to be projected close ($\|p_u - q_i\|$ is small) to that user in the latent space.

3 Proposed Methods

In this section, we propose a series of novel methods. First, due to the difference of individuality, explainability matrix is generated based on similarity, getting the explainable matrix factorization based on similarity. Second, considering the neighborhood information in the latent space, similar users should be close to this explainable item (this user should also be close to the similar explainable items), so the neighbors are added to the explainability constraint term. Further, considering the difference of neighbors, the weights of neighbors are added to better exploit the whole relationship.

3.1 Similarity-Based Explainable Matrix Factorization (SEMF)

It is more reasonable to use similarity between users and users or items and items than to use the empirical conditional probability of rating. The similarity reflects

the strength of relationship of users (items). If users have close relationship, the influence of neighbors on user should be larger. Then the high rating items of users who are similar to active user are also more explainable. Therefore, the explainability matrix W_s is defined as Eqs. 4 and 5. Using the W_s in the EMF, the SEMF is given as follows:

$$J = \sum_{u,i \in R} (r_{u,i} - p_u q_i^T)^2 + \frac{\beta}{2} (\|p_u\|^2 + \|q_i\|^2) + \frac{\gamma}{2} \|p_u - q_i\|^2 W_s(u, i), \quad (7)$$

where $W_s(u, i)$ is obtained by the similarity Pearson Correlation Coefficient; the $\frac{\beta}{2} (\|p_u\|^2 + \|q_i\|^2)$ represents the regularization term adjusted by coefficient β , which can prevent over fitting, and γ denotes the coefficient of explainability term which controls the smoothness of the model and tradeoff between explanation and accuracy. To minimize the loss function J to achieve prediction, many methods such as machine learning including gradient descent [9, 10, 12] and alternating least squares [4, 8] are applied. In this study, the stochastic gradient descent is adopted in which each sample is updated iteratively, rapidly descending to approach to a minimum. The method is given as follows:

$$\begin{aligned} p_u^{(t+1)} &= p_u^t + \alpha(2(r_{u,i} - p_u q_i^T)q_i - \beta p_u - \gamma(p_u - q_i)W_s(u, i)), \\ q_i^{(t+1)} &= q_i^t + \alpha(2(r_{u,i} - p_u q_i^T)p_u - \beta q_i + \gamma(p_u - q_i)W_s(u, i)). \end{aligned} \quad (8)$$

3.2 Neighborhood-Based Explainable Matrix Factorization (NEMF)

The neighborhood-based CF makes use of neighborhood information in the original space to achieve explanation. In the subsection we make use of the neighborhood information in the latent space to improve accuracy and explanation. Because the items with high explainable relationship to user are close to this active user in the latent space. Similarly, if we add the neighborhood information in the latent space to the explainability term, the items should also be close to the neighbors of this active user in the latent space. For the neighbors of items, it is analogous. Hence we can get the NEMF approach as follows:

User - Based :

$$J = \sum_{u,i \in R} [(r_{u,i} - p_u q_i^T)^2 + \frac{\beta}{2} (\|p_u\|^2 + \|q_i\|^2) + \frac{\gamma}{2} \sum_{v \in neighbors} \|p_v - q_i\|^2 W_{u_s}(u, i)],$$

Item - Based :

$$J = \sum_{u,i \in R} [(r_{u,i} - p_u q_i^T)^2 + \frac{\beta}{2} (\|p_u\|^2 + \|q_i\|^2) + \frac{\gamma}{2} \sum_{j \in neighbors} \|p_u - q_j\|^2 W_{i_s}(u, i)], \quad (9)$$

where *neighbors* is the neighbors of this active user (item) including itself. When the item is explainable to the active user, meaning $W_s(u, i) \geq \theta$. Then the user and the item are close in the latent space, i.e., $\|p_u - q_i\|$ is close to zero. The greater $W_s(u, i)$ is, the closer to zero $\|p_u - q_i\|$ is. Adding the neighbors' effect can balance the relationships of users (items), which is more reasonable to achieve

explanation. We apply statistic gradient descent to minimize the objective function. Taking the user-based for example, the updates of $p_u^{(t+1)}$ and $q_i^{(t+1)}$ are represented as follows:

$$\begin{aligned} p_u^{(t+1)} &= p_u^t + \alpha(2(r_{u,i} - p_u q_i^T)q_i - \beta p_u - \gamma(p_u - q_i)W u_s(u, i)), \\ q_i^{(t+1)} &= q_i^t + \alpha(2(r_{u,i} - p_u q_i^T)p_u - \beta q_i + \sum_{v \in neighbors} \gamma(p_v - q_i)W u_s(u, i)). \end{aligned} \quad (10)$$

3.3 NEMF Adding Neighbors' Weight (NEMF++)

Further, considering that the relationships of neighbors are different, the similarity of users (or items) is added into the explainability term to represent the different relationships. In the above subsection, we make the user and his neighbors close to the item in latent space to improve the performance. In this part, the similarities of neighbors are introduced. Because the neighbors are close to the same one, the neighbors are close each other in the latent space. Introducing the similarities of neighbors make the degree of proximity different. The approach makes explanation full. The equation is given as follows:

User - Based :

$$J = \sum_{u,i \in R} [(r_{u,i} - p_u q_i^T)^2 + \frac{\beta}{2}(\|p_u\|^2 + \|q_i\|^2) + \frac{\gamma}{2} \sum_{v \in neighbors} \|p_v - q_i\|^2 W u_s(u, i) sim_{u,v}],$$

Item - Based :

$$J = \sum_{u,i \in R} [(r_{u,i} - p_u q_i^T)^2 + \frac{\beta}{2}(\|p_u\|^2 + \|q_i\|^2) + \frac{\gamma}{2} \sum_{j \in neighbors} \|p_u - q_j\|^2 W i_s(u, i) sim_{i,j}]. \quad (11)$$

Anyhow, adding neighbors to explainability term, makes similar users and the explainable item are all close to active user in the latent space. The relationships of neighbors in the latent space prominently improve the model performance.

4 Experiments

4.1 Experiment Prepared

In our experiment, we use the public data set-100k MovieLens provided by GroupLens [7] which contains 10^5 ratings on scale of 1 to 5 from 943 users on 1862 movies. We split data into two parts. Namely the 10% of the ratings from each user are regarded as the test set, and the remaining ratings are the training set. The data is normalized from 0 to 1 in our experiment and the parameters are often tuned by cross validation.

We compare our approaches including the SEMF, NEMF and NEMF++ with the best baseline EMF which represents the state-of-art performance on the explainable recommendation task without any additional data source. The recommendation list is generated by top-n recommendation method, which includes the highest n ratings items.

4.2 Evaluation Metrics

For the top-n recommendation, two of the most common metrics are applied to evaluate the performances of proposed approaches, the same as, for the explanation, two of the metrics are applied. The metrics are given as follows:

Mean Average Precision (MAP) [11]: MAP evenly measures the ratio of the number of the test list contained in the top-n recommendation list for an $user_{u,i}$. The MAP is the mean of the average precision for each user, which can be given by

$$MAP@n = \frac{\sum_{u=1}^{N_u} AP(n)}{N_u}, \quad (12)$$

where N_u is the number of the users. And AP can be written as:

$$AP@n = \frac{\sum_{k=1}^n P(k)}{\min(l, n)}, \quad (13)$$

where $P(k)$ means the precision at cut-off k in the item list, and l is the length of actual test data list.

Area Under ROC Curve (AUC) [5]: AUC indicates the ability of distinguishing the relevant objects from the irrelevant objects, which is defined as:

$$AUC = \frac{n_1 + n_2 * 0.5}{N}, \quad (14)$$

where N is a large number for N independent comparisons. When we randomly select a rated item and an unrated item, if the rating of rated item is larger than the other, n_1 will add one. If they are equal, n_2 will add one.

Mean Explainability Precision (MEP) [1]: MEP measures the explainable precision for items recommended to this active user. It is defined as:

$$MEP = \frac{1}{U} \sum_{u=1}^U \frac{N_{exp}}{L}, \quad (15)$$

where N_{exp} is the number of explainable items (if $E_{u,i} \geq \theta$, we think the item is explainable) in the top-n recommendation list and L is the number of recommended (top-n) items for each user.

Mean Explainability Recall (MER) [1]: MER measures the explainable recall defined as:

$$MER = \frac{1}{U} \sum_{u=1}^U \frac{N_{exp}}{E}, \quad (16)$$

where the E is the number of all explainable items for an user. In most cases, the MEP and MER are contrary, so it is difficult to evaluate the performance. Hence we introduce the F-measure.

F-Measure [13]: F-measure is the harmonic average of the precision and recall, which can be used to evaluate proposed approaches. The F_1 is defined as:

$$F_1 = \frac{2 * MEP * MER}{MEP + MER}. \quad (17)$$

4.3 Simulation and Discussion

In the simulation, the dimension f and the number of neighbors N is varied. Other parameters are found by cross-validation: $\alpha = 0.001, \beta = 0.01, \theta = 0.01, \gamma = 0.05$. γ is not set very small, because explanation is important for us and we compare the explanation with many approaches. If the γ is very small, the difference of methods is also small. For the AUC, we set the frequency of comparison as 10^6 . The recommendation list length L is 50. We compare our three approaches with the baseline state-of-the-art method Explanation Matrix Factorization (EMF). The simulation of the four methods for explanation is as follows:

Table 1. Compare MAP@50: Top row: varying the dimension of latent space f , $N = 50$. Bottom row: varying the number of neighbors N , $f = 10$

MAP@50								
f	EMF		SEMF		NEMF		NEMF++	
	<i>User</i>	<i>Item</i>	<i>User</i>	<i>Item</i>	<i>User</i>	<i>Item</i>	<i>User</i>	<i>Item</i>
5	0.01484	0.01558	0.01485	0.01559	0.01477	0.01599	0.01559	0.01590
10	0.01538	0.01611	0.01536	0.01613	0.01526	0.01667	0.01613	0.01672
20	0.01587	0.01745	0.01592	0.01750	0.01593	0.01835	0.01792	0.01845
50	0.01620	0.01688	0.01622	0.01698	0.01603	0.01785	0.01733	0.01795
N								
5	0.01513	0.01613	0.01517	0.01620	0.01540	0.01670	0.01578	0.01667
10	0.01512	0.01606	0.01516	0.01617	0.01529	0.01665	0.01590	0.01674
20	0.01518	0.01607	0.01537	0.01612	0.01527	0.01672	0.01604	0.01672
50	0.01538	0.01611	0.01536	0.01613	0.01526	0.01667	0.01613	0.01672

Table 2. Compare AUC: Top row: varying the dimension of latent space f , $N = 50$. Bottom row: varying the number of neighbors N , $f = 10$

AUC								
f	EMF		SEMF		NEMF		NEMF++	
	<i>User</i>	<i>Item</i>	<i>User</i>	<i>Item</i>	<i>User</i>	<i>Item</i>	<i>User</i>	<i>Item</i>
5	0.8136	0.8153	0.8136	0.8154	0.8139	0.8163	0.8163	0.8167
10	0.8168	0.8185	0.8169	0.8187	0.8176	0.8199	0.8201	0.8205
20	0.8277	0.8291	0.8277	0.8292	0.8287	0.8301	0.8307	0.8310
50	0.8477	0.8488	0.8477	0.8489	0.8486	0.8495	0.8505	0.8506
N								
5	0.8167	0.8185	0.8170	0.8193	0.8177	0.8201	0.8193	0.8204
10	0.8167	0.8183	0.8169	0.8190	0.8177	0.8200	0.8197	0.8204
20	0.8167	0.8183	0.8168	0.8188	0.8176	0.8199	0.8198	0.8205
50	0.8168	0.8185	0.8169	0.8187	0.8176	0.8199	0.8201	0.8205

Observing Tables 1 and 2 on the whole, we can find that all methods based on item are better than the corresponding methods based on user in the top-n recommendation. Because there are more data in the front of the data set, which contributes better similarity based on item. Besides, with the increasing of f , the performance becomes better. However, the change is small when varying N . The dimension of latent space represents the feature. With the increasing of f , the features can better represent the users (items). From the experimental result, it can be seen that the change is small when varying N . There are minor influence to the result with the change of the number of neighbors.

In the detail, the performance of proposed methods is gradually improved. The proposed methods are better on MAP and AUC with the change of dimension f or number of neighbors N than the baseline method EMF in most cases. And the NEMF++ greatly outperforms the benchmark methods EMF, suggesting that considering neighborhood information can improve the accuracy performance in the top-n recommendation.

In terms of explanation, the comparisons are given in Tables 3, 4 and 5. Similarly, on the whole, all methods based on user are better than the corresponding methods based on item in the terms of explanation, and this is greatly different from MAP and AUC. The result is caused by the sparse data set. There are sometimes no rating for items, especially at the back of data set. If there are no rating for the items, the explanation of this item will be poor. If the threshold is fixed, the number of explanation will become small, so the MEP of methods based on item is smaller than the one based on user. Because we make use of the relationship of users with users, users with items and items with items to achieve the goal of explanation, for the MEP and MER, there is minor influence with the change of dimension f , but there is greater influence with the change of number N of neighbors. In the same threshold case, when the number N of neighbors

Table 3. Compare MEP@50: Top row: varying the dimension of latent space f , $N = 50$. Bottom row: varying the number of neighbors N , $f = 10$

MEP@50								
f	EMF		SEMF		NEMF		NEMF++	
	<i>User</i>	<i>Item</i>	<i>User</i>	<i>Item</i>	<i>User</i>	<i>Item</i>	<i>User</i>	<i>Item</i>
5	0.9339	0.6164	0.9265	0.6022	0.9274	0.6062	0.9311	0.6075
10	0.9278	0.6093	0.9200	0.5931	0.9205	0.5982	0.9255	0.5977
20	0.9215	0.6182	0.9128	0.6038	0.9140	0.6102	0.9191	0.6120
50	0.9179	0.6324	0.9087	0.6169	0.9106	0.6244	0.9175	0.6257
N								
5	0.4124	0.1889	0.4129	0.1897	0.4137	0.1920	0.4177	0.1919
10	0.5842	0.3027	0.5843	0.3035	0.5855	0.3066	0.5911	0.3069
20	0.7538	0.4337	0.7547	0.4347	0.7556	0.4397	0.7618	0.4398
50	0.9278	0.6093	0.9200	0.5931	0.9205	0.5982	0.9255	0.5977

Table 4. Compare MER@50: Top row: varying the dimension of latent space f , $N = 50$. Bottom row: varying the number of neighbors N , $f = 10$

MER@50								
f	EMF		SEMF		NEMF		NEMF++	
	<i>User</i>	<i>Item</i>	<i>User</i>	<i>Item</i>	<i>User</i>	<i>Item</i>	<i>User</i>	<i>Item</i>
5	0.06816	0.05305	0.07197	0.05221	0.07206	0.05729	0.07249	0.05763
10	0.06759	0.05200	0.07131	0.05090	0.07137	0.05598	0.07193	0.05607
20	0.06703	0.05322	0.07062	0.05236	0.07075	0.05790	0.07131	0.05828
50	0.06663	0.05481	0.07013	0.05376	0.07034	0.05940	0.07109	0.05966
N								
5	0.11585	0.10380	0.11603	0.06728	0.11672	0.10663	0.11885	0.10676
10	0.09965	0.08705	0.09975	0.06714	0.10022	0.08908	0.10196	0.08933
20	0.08643	0.07103	0.08568	0.06101	0.08591	0.07235	0.08712	0.07256
50	0.06759	0.05200	0.07131	0.05090	0.07137	0.05598	0.07193	0.05607

becomes larger, the explainable items become more. So the MEP becomes large with the increasing of number N of neighbors. But the MER usually is opposite. So we use the parameter F_1 to measure the two metrics.

In the detail, we compare the proposed methods with the baseline method. It can be seen that the MEP of proposed methods are slightly poorer than EMF, but the MER of proposed methods are significantly better than EMF. It suggests that we can recommend more explainable items from all explainable items to this active user. Meanwhile we can clearly find that NEMF++ is optimal in the balanced metric F_1 .

Table 5. Compare F_1 : Top row: varying the dimension of latent space f , $N = 50$. Bottom row: varying the number of neighbors N , $f = 10$

F_1								
f	EMF		SEMF		NEMF		NEMF++	
	<i>User</i>	<i>Item</i>	<i>User</i>	<i>Item</i>	<i>User</i>	<i>Item</i>	<i>User</i>	<i>Item</i>
5	0.1270	0.0977	0.1336	0.0961	0.1337	0.1047	0.1345	0.1053
10	0.1260	0.0958	0.1324	0.0937	0.1325	0.1024	0.1335	0.1025
20	0.1250	0.0980	0.1311	0.0964	0.1313	0.1058	0.1324	0.1064
50	0.1242	0.1009	0.1302	0.0989	0.1306	0.1085	0.1319	0.1089
N								
5	0.1809	0.1340	0.1812	0.0993	0.1821	0.1371	0.1850	0.1372
10	0.1703	0.1352	0.1704	0.1100	0.1711	0.1381	0.1739	0.1384
20	0.1551	0.1221	0.1539	0.1070	0.1543	0.1243	0.1564	0.1246
50	0.1260	0.0958	0.1324	0.0937	0.1325	0.1024	0.1335	0.1025

5 Conclusion

From the experimental result, it can be seen that we can recommend more explainable items to a customer and keep high accuracy. If we use our methods to recommend item A to a customer, We can explain to this customer that “for the recommended item A, your friends bought it. The maximum of similarity among them is B. They comprehensively gave rating C to this item.” Then recommendation achieves explanation. In this study, we change the calculation method of explainability matrix and propose a novel explainable matrix factorization, which jointly characterizes both neighborhood information and user-item interactions in the latent space to achieve better explanation. The experimental results show that our methods outperform the baseline method in most cases. In the future, we hope to extend our methods into other recommendation fields and propose some metrics to evaluate the explainability further.

Acknowledgement. This work was supported by the National Natural Science Foundation of China (No. 61602048) and the Fundamental Research Funds for the Central Universities (No. NST20170206).

References

1. Abdollahi, B., Nasraoui, O.: Explainable matrix factorization for collaborative filtering. In: International Conference Companion on World Wide Web, pp. 5–6 (2016)
2. Abdollahi, B., Nasraoui, O.: Using explainability for constrained matrix factorization. In: The Eleventh ACM Conference, pp. 79–83 (2017)
3. Amatriain, X.: Past, present, and future of recommender systems: an industry perspective. In: International Conference on Intelligent User Interfaces, p. 1 (2016)

4. Bell, R.M., Koren, Y.: Scalable collaborative filtering with jointly derived neighborhood interpolation weights. In: IEEE International Conference on Data Mining, pp. 43–52 (2007)
5. Davis, J., Goadrich, M.: The relationship between precision-recall and ROC curves. In: International Conference on Machine Learning, pp. 233–240 (2006)
6. Gedikli, F., Jannach, D., Ge, M.: How should i explain? A comparison of different explanation types for recommender systems. *Int. J. Hum. Comput. Stud.* **72**(4), 367–382 (2014)
7. Harper, F.M., Konstan, J.A.: The MovieLens datasets: history and context. *ACM Trans. Interact. Intell. Syst. (TiiS)* **5**(4) (2016). Article No. 19
8. Hu, Y., Koren, Y., Volinsky, C.: Collaborative filtering for implicit feedback datasets. In: Eighth IEEE International Conference on Data Mining, pp. 263–272 (2009)
9. Koren, Y.: Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 426–434 (2008)
10. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. *Computer* **42**(8), 30–37 (2009)
11. Manning, C.D., Raghavan, P., Schtze, H.: Chapter 8: Evaluation in Information Retrieval. Cambridge University Press, New York (2009)
12. Paterek, A.: Improving regularized singular value decomposition for collaborative filtering. In: Proceedings of KDD Cup Workshop (2007)
13. Powers, D.M.W.: Evaluation: From precision, recall and f-measure to ROC, informedness, markedness correlation. *J. Mach. Learn. Technol.* **2**(1), 37C63 (2011)
14. Vig, J., Sen, S., Riedl, J.: Tagsplanations: explaining recommendations using tags. In: International Conference on Intelligent User Interfaces, pp. 47–56 (2009)
15. Zhang, Y., Lai, G., Zhang, M., Zhang, Y., Liu, Y., Ma, S.: Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. *ACM* (2014)