A Rating Bias Formulation based on Fuzzy Set for Recommendation

1st Mingming Li[†,‡], 2nd Fuqing Zhu[*,†], 3rd Jiao Dai[†], 4th Liangjun Zang[†], 5th Yipeng Su[†], 6th Jizhong Han[†], 7th Songlin Hu[†,‡]
[†] *Institute of Information Engineering, Chinese Academy of Sciences*
[‡] *School of Cyber Security, University of Chinese Academy of Sciences*
Beijing, China
{limingming,zhufuqing,daijiao,zangliangjun,suyipeng,hanjizhong,husonglin}@iie.ac.cn

Abstract—**In recommender systems, the user uncertain preference results in unexpected ratings. Previous approaches (e.g., BiasMF) only adjust the rating value based on the bias vector, ignoring the uncertainty of rating. This paper makes an initial attempt in integrating the influence of user uncertain degree and user rating bias into the matrix factorization framework, simultaneously. An approach based on fuzzy set, called fuZzy Matrix Factorization (ZMF), is proposed. Specifically, a fuzzy set of *like* is defined for each user, and the membership function is utilized to measure the degree of an item belonging to the fuzzy set. Then, the user uncertain preference matrix is obtained, which could explain and represent the user bias and uncertainty effectively. Furthermore, to enhance the computational impact on sparse matrix, the uncertain preference is formulated as a side-information for fusion. Besides, the proposed approach could be extended to others due to independency on additional data sources. Experimental results on three datasets show that ZMF produces an effective improvement.**

Keywords-**Recommender Systems, Fuzzy Set, Uncertain Preference, Rating Bias**

I. Introduction

With the growing number of products and services, recommender systems have become necessary tools to discover information of interest for users. Many recommendation approaches have been proposed and made breakthroughs in various applications. Matrix Factorization (MF) [1] is one of the most famous and successful approaches, which learns the discriminative latent factors for users and items by factorizing user-item interaction matrix. However, there still exist two drawbacks of the above approaches: 1) The rating information is uncertain and imprecise. Different users have various evaluation criteria, and different ratings also have different degrees of preferences. 2) The rating matrix is always sparse in the realistic world, which may compromise the performance of recommender systems.

To handle the uncertain and imprecise information of rating, the user/item bias vector is utilized to adjust original ratings, such as BiasMF [1]. Nevertheless, the impact of uncertain ratings is neglected. Specifically, as shown in Figure 1, items with very low/high ratings explicitly indicate that users really dislike/like them, whereas the majority of intermediate ratings are uncertain. The same rating may represent different meanings since the evaluation concepts of users are different. For an item with rating = 4, it is uncertain to determine the preference of user i, while it indicates that the user j dislikes this item clearly. For an item rating = 7 rated by users i and j, we are not sure whether they like it. But the degree of preference is measurable that user i may be more prone to like this item than user j. To our best knowledge, previous approaches based on a crisp data model fail to capture the notions of uncertainty. A sharp boundary is often defined to discriminate members belonging to the set of ***like*** (e.g., rating>4) from non-members. Such approaches have a drawback: it is rough to choose a global boundary value for all users. To overcome above limitations, we represent the notion of ***like*** for each user by fuzzy set theory [2], [3]. First, we divide the original rating into three groups (i.e., *dislike,uncertain,like*), according to the user rating bias. *Note: the same rating may correspond to different users' groups.* Then, we calculate the possibility of each rating which belongs to the fuzzy set ***like***, avoiding the limitations of the sharp boundary.

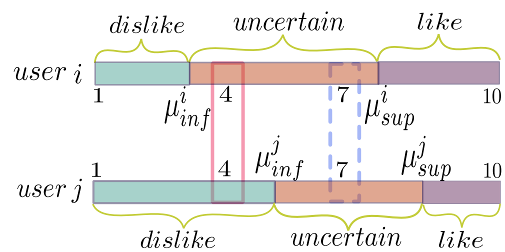


Fig. 1: Illustration of rating bias

To alleviate the sparsity problem, existing approaches often construct a hybrid model by combining the auxiliary information to MF, such as Collaborative Topic Regression (CTR) [4], Collaborative Deep Learning (CDL) [5] and Convolutional Matrix Factorization (ConvMF) [6]. Above approaches provide a new perspective to

*Corresponding author

alleviate the sparsity problem and achieve significant improvement. Nevertheless, further optimization is not effective due to the limitation of the data source. Thus, the view of reconstructing information from original data is proposed. In [7]–[9], the item co-occurrence matrix is utilized as side-information. Our work is motivated by exploring some potential information from the original information without any additional data source.

Given the above two considerations, an approach based on fuzzy set, called fuZzy Matrix Factorization (ZMF), is proposed. More specifically, ZMF first explores the user uncertain preference information by using fuzzy set, and then make them as the side-information for jointly matrix factorization. The user uncertain preference matrix of ZMF is constructed according to the user-item interaction information, where each element indicates the degree of one item belonging to a user ***like*** set. Moreover, we extend our *user uncertain preference information* strategy to ConvMF (i.e., ZMF-C). Experiments show that our proposed approach consistently achieve a stable improvement on three real-world datasets.

The main contributions of this paper are summarized as follows:

- User uncertain preference is introduced to eliminate the effect of fuzzy information. To the best of our knowledge, this is an initial attempt to measure the user bias by fuzzy set theory in recommender systems.
- Jointly matrix factorization based on fuzzy set (ZMF) is proposed to integrate user uncertain preference and rating information, which could remit the uncertain ratings and data sparsity problems.
- On three public explicit feedback datasets, extensive experimental results demonstrate that our proposed approach produces competitive performances from multiple perspectives.

II. Related Work

This paper focuses on modeling user uncertain preference by fuzzy sets based on matrix factorization in recommender systems. Thus, we will first discuss the related work of matrix factorization, and then introduce the fuzzy tools in recommender systems.

A. Matrix Factorization

Probabilistic Matrix Factorization [10] models the user-item rating matrix as a product of two lower-rank user and item matrices. The fact that much of the observed variation in rating values is due to effects associated with either users or items, known as biases or intercepts. Thus, BiasMF breaks down the observed rating into four components, i.e., global average, item bias, user bias, and user-item interaction. To alleviate data sparsity problem and cold-start problem, many hybrid models have been proposed by combining the auxiliary information (review comments, social relationships, etc.) to the PMF. Convolutional Matrix Factorization (ConvMF) [6] is the representative work, which mixes PMF and CNN [11]. ConvMF can learn an efficient representation for item with the item side-information by utilizing the network of CNN due to the advantages of mining the local features. ConvMF makes superior improvement to PMF, and it shows the significant impact of the side-information. Our work is also a jointly matrix factorization. But it is different from all the above works, our work focuses on handling the rating bias and uncertain degree only depending on the original rating information rather than additional data sources.

B. Fuzzy Tools in Recommender Systems

Fuzzy tools [12] are beneficial for improving the performance of recommender systems. The item features and user feedback are often subjective, imprecise and uncertain in real-world applications. In this case, the fuzzy linguistic approaches [13], [14] could be utilized to address the vague text information. Besides, the fuzzy representation and fuzzy similarity metrics approaches [15], [16] are presented by fuzzy set theory and focus on the text information in Content-Based recommender systems. In the early days, most of above fuzzy based approaches are focused on modeling the attributes of items, instead of the preferences for each user. However, the preferences are the main source of uncertainty in recommender systems. In the following study, the collaborative filtering approaches with fuzzy tools [17] is proposed to make full use of the preference values without any additional information, from which the rating information is formulated by global map function without considering the user-specific bias. In this paper, both the rating bias and uncertainty are comprehensively considered by fuzzy set theory, aiming to explain and represent the original imperfect rating information.

III. Proposed Approach

As aforesaid, above approaches have less space for improvement due to a limited data source, meanwhile, they could not capture the uncertain preference of users. To address this problem, we exploit the user uncertain preference information from the original data, and then make it as the side-information for modeling. In this section, we first introduce the fuzzy set theory. Then, the construction of user uncertain preference matrix is described. After that, the jointly matrix factorization approach ZMF is proposed to realize the integration of different information. Finally, we will show our advancement that an expansion to ConvMF by employing the user uncertain preference for more comprehensive recommendation.

Suppose there are N users, M items, and a user-item rating matrix $R \in \mathcal{R}^{N \times M}$. Let $R_{i,j}$ represents the rating of user i on item j; $U \in \mathcal{R}^{N \times K}$ and $V \in \mathcal{R}^{M \times K}$ be latent user and item feature matrices, with column vectors U_i

and V_j representing user-specific and item-specific latent feature vectors respectively, where $K \ll min(M, N)$.

A. Fuzzy Set Theory

To deal with fuzzy information, [18] presents fuzzy set theory, which has been proved as a successful technique for modeling the fuzzy information in many areas such as publish/subscribe system [19], semantic web [3], [20] and database [21].

Definition [2]: Let $\mathcal{U}$ be a universe of discourse. A fuzzy value on $\mathcal{U}$ is characterized by a fuzzy set $\widetilde{F}$ in $\mathcal{U}$. A membership function

$$\psi_{\widetilde{F}} : \mathcal{U} \rightarrow [0, 1] \tag{1}$$

is defined for the fuzzy set $\widetilde{F}$, where $\psi_{\widetilde{F}}(u)$ for each $u \in \mathcal{U}$, denotes the degree of membership of u in the fuzzy set $\widetilde{F}$. For example, $\psi_{\widetilde{F}}(u) = 0.5$ means that u is "likely" to be an element of $\widetilde{F}$ by a degree of 0.5. As the generalization of the characteristic function in classical mathematics set theory, the membership function allows to express gradual set membership, which can eliminate the adverse effect of the sharp boundary. The membership function can be defined according to specific problems, in which the most popular approaches are normal distribution curve and S-type distribution.

B. User Uncertain Preference Matrix

Users have different evaluation criteria, and different ratings have different degrees of preferences. Thus, we handle the rating bias by the fuzzy set. In this paper, we define a specific fuzzy set of the concept of ***like*** for user i, denoted as $\widetilde{F}_i$, $\widetilde{F} = \{\widetilde{F}_i | i = 1, \ldots, N\}$. We use the S-type function to metric the membership degrees of $\widetilde{F}_i$:

$$\psi_{\widetilde{F}_i}(R_{i,j}) = \begin{cases} 0 & R_{i,j} \leq \mu^i_{inf} \\ \frac{1+\xi_i}{1+a(R_{i,j}-\mu^i_{inf})^b} & \mu^i_{inf} < R_{i,j} < \mu^i_{sup} \\ 1 & R_{i,j} \geq \mu^i_{sup} \end{cases} \tag{2}$$

where $a > 0, b < 0, \xi_i > 0$, and μ^i_{inf} and μ^i_{sup} is the infimum and supremum of user i, respectively. As shown in Figure 2, the <infimum, supremum> of user i and j is $< \mu^i_{inf}, \mu^i_{sup} >$ and $< \mu^j_{inf}, \mu^j_{sup} >$, respectively. For an item with a rating equal to r_1, the user i explicitly dislikes it because of $r_1 < \mu^i_{inf}$, while user j likes it with a possibility of 0.4. When the rating of an item is equal to r_2, we could know that both users may like this item with larger probability.

Selection of a and b: For the uncertain member, the possibility is formulated as the scope of [0, 1]. The membership function is to map the original score to this continuous interval. The higher possibility denotes the larger original score. Therefore, the membership function should satisfy the following conditions:

(i) the membership function should be monotonic non-decreasing;

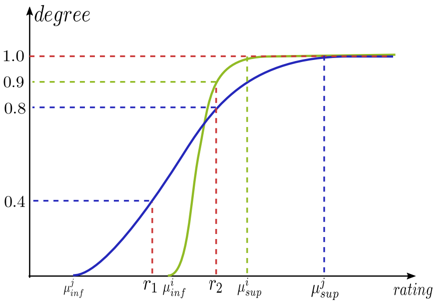


Fig. 2: The membership function representation for user ***like***

(ii) the membership function must be continuous (smooth) and gradual (non-saturated);

(iii) the lower bound of membership function value tends to 0, while the upper bound tends to 1.

First, for the condition (i), we set $a > 0, b < 0$. Second, for the condition (ii) (Especially, the function cannot be increased too violently.), we discuss the value of a and b, respectively. For parameter b, the function tends to be more easily saturated when b is much smaller. Thus, b cannot be too small. For parameter a, it determines the length of the gentle gradient interval. Because the uncertainty score interval is different for each user. The selection of the parameter a should be related to the infimum and supremum for each user. Finally, for the condition (iii), we set an additional variable epsilon. Moreover, the value of $< \mu^i_{inf}, \mu^i_{sup} >$ is set based on the experimental verification, which is analysed in the corresponding experimental section. In summary, we choose S-type function as the membership function, from which the parameters $a = \mu^i_{sup} - \mu^i_{inf} + 1$,$b = -3$, and $\xi_i = a(\mu^i_{sup} - \mu^i_{inf})^b$ for our proposed approach. In this way, it can be realized a trade-off between the nature of fuzzy set theory and the actual scene, simultaneously.

According to above map function, we can construct a user uncertain preference matrix $S \in \mathcal{R}^{N \times M}$, which indicates the degree of user preference (***like***) on each item. And each entry of S can be computed as:

$$S_{i,j} = \psi_{\widetilde{F}_i}(R_{i,j}) \tag{3}$$

C. ZMF

To integrate the user uncertain preference information into MF framework, we consider to jointly decompose the user-item interaction matrix and the user uncertain preference matrix. However, there are two problems: rating information is not in the same space as the user uncertain preference information (the former is in the real number space, and the latter is in the probability space); the user latent factors in the user uncertain preference information represent the possibility of user preference, and thus we should ensure that it is non-

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

negative. To address above two problems, we present an approach of joint orthogonal non-negative matrix factorization that measures the latent factors of users in Public spaces (ZMF). Essentially, this approach is to obtain a more accurate user potential representation vector by increasing the user regular constraints. This approach refers the technology of non-negative matrix factorization [22] and orthogonal non-negative matrix factorization [23]. The objective function of ZMF can be expressed as follows:

$$\begin{aligned}\min_{U,V,P,Q,D} \mathcal{L} =& ||W \odot (R - UV^T)||_F^2 + ||W \odot (S - PQ^T)||_F^2 \\ &+ ||U - D||_F^2 + ||P - D||_F^2 + \lambda_U ||U||_F^2 \\ &+ \lambda_V ||V||_F^2 + \lambda_Q ||Q||_F^2 + \lambda_P ||P||_F^2 \\ &s.t.\ D^T D = I, D \geq 0, U \geq 0, P \geq 0 \end{aligned} \tag{4}$$

where U and P respectively represent the user characteristics in the rating information and preference information; D is the public space shared by U and P. What's more, in order to ensure the uniqueness of the space, we force to D to satisfy the orthogonal conditions. The problem in Eq.(4) could be solved by the block-coordinate descent algorithm [24], due to the same status of them and separating easily.

Optimize U**.** By fixing V, P, Q, D, the optimization problem becomes:

$$\min_{U \geq 0} ||W \odot (R - UV^T)||_F^2 + ||U - D||_F^2 + \lambda_U ||U||_F^2 \tag{5}$$

Following the standard theory of constrained optimization, the Lagrangian function can be defined as follows:

$$\begin{aligned}\min_U \mathcal{L}(U) =& ||W \odot (R - UV^T)||_F^2 + ||U - D||_F^2 \\ &+ \lambda_U ||U||_F^2 - Tr(\Gamma U^T)\end{aligned} \tag{6}$$

where $\Gamma \in \mathcal{R}^{N \times K}$ is the Lagrangian multiplier. We take the gradient $\nabla_U \mathcal{L}(U) = 0$, and use the Karush-Kuhn-Tucker (KKT) complementarity condition $\Gamma_{i,k} U_{i,k} = 0$, we can obtain:

$$\begin{aligned}&[-R_i W_i V + U_i V^T W_i V + (1 + \lambda_U) U_i - D_i]_k U_{i,k} = 0 \\ &i \in \{1, \ldots, N\}, k \in \{1, \ldots, K\}\end{aligned} \tag{7}$$

where $W_i \in \mathcal{R}^{M \times M}$ is a diagonal matrix with $W_{i,j}$ as its diagonal element, R_i is a vector with $R_{i,j}$ for user i, S_i is a vector with $S_{i,j}$ for user i. Let $(X_{i,k})^+ = (|X_{i,k}| + X_{i,k})/2$, $(X_{i,k})^- = (|X_{i,k}| - X_{i,k})/2$, and $X = X^+ - X^-$. So the Eq.(7) can be writed as:

$$\begin{aligned}\{-[(R_i W_i V)^+ + D_i + U_i (V^T W_i V)^-] + [(R_i W_i V)^- \\ + U_i (V^T W_i V)^+ + (1 + \lambda_U) U_i]\}_k U_{i,k} = 0\end{aligned} \tag{8}$$

According to the [22], the update rule for U_i is:

Algorithm 1: ZMF

Input: $R, W, \lambda_U, \lambda_V, \lambda_P, \lambda_Q$
Output: U, V

```
Construct user uncertain preference matrix S by Eq.(3);
Initialize U, V, P, Q, D, Ψ by uniform(0,1);
repeat
    for i ∈ {1, ..., N} do
        Update U_i using Eq.(9);
        Update P_i using Eq.(10);
        D_i ← D_i ⊙ sqrt( (U_i + P_i + (D_i Ψ)^-) / (2D_i + (D_i Ψ)^+) );
    end
    for j ∈ {1, ..., M} do
        Update V_j using Eq.(11);
        Update Q_j using Eq.(12);
    end
until L is convergent;
```

$$U_i \leftarrow U_i \odot \sqrt{\frac{(R_i W_i V)^+ + D_i + U_i (V^T W_i V)^-}{(R_i W_i V)^- + U_i (V^T W_i V)^+ + (1 + \lambda_U) U_i}} \tag{9}$$

Optimize P**.** Similar to U_i, the update rule for P_i can be obtained as follows:

$$P_i \leftarrow P_i \odot \sqrt{\frac{(S_i W_i Q)^+ + D_i + P_i (Q^T W_i Q)^-}{(S_i W_i Q)^- + P_i (Q^T W_i Q)^+ + (1 + \lambda_P) P_i}} \tag{10}$$

Optimize V**,**Q**.** Similarly, setting the gradient $\nabla_V \mathcal{L}(V) = 0$, and $\nabla_Q \mathcal{L}(Q) = 0$, respectively, we can get their update rule as:

$$V_j \leftarrow (R_j W_j U)(U^T W_j U + \lambda_V E_K)^{-1} \tag{11}$$

$$Q_j \leftarrow (S_j W_j P)(P^T W_j P + \lambda_Q E_K)^{-1} \tag{12}$$

where the definitions of W_j, R_j and S_j are similar to W_i, R_i and S_i, respectively.

Optimize D**.** By fixing U, V, P, Q, the optimization problem becomes:

$$\begin{aligned}&\min_{D \geq 0} ||U - D||_F^2 + ||P - D||_F^2 \\ &s.t.\ D^T D = I\end{aligned} \tag{13}$$

Because Eq.(13) contains the equality and inequality constraints simultaneously, we use Lagrangian multipliers and KKT condition to solve this problem. The Lagrangian function is:

$$\begin{aligned}\min_D \mathcal{L}(D) =& ||U - D||_F^2 + ||P - D||_F^2 \\ &+ Tr(\Psi (D^T D - I)) - Tr(\Phi D^T)\end{aligned} \tag{14}$$

where $\Phi \in \mathcal{R}^{N\times K}$, $\Psi \in \mathcal{R}^{K\times K}$(a symmetric matrix). Making the gradient $\nabla_D \mathcal{L}(D) = 0$ and using the KKT condition $\Phi_{i,k} D_{i,k} = 0$, we can obtain:

$$(-U - P + 2D + D\Psi)_{i.k} D_{i,k} = 0 \quad (15)$$

$$\{-[U + P + (D\Psi)^-] + [2D + (D\Psi)^+]\}_{i.k} D_{i,k} = 0 \quad (16)$$

where $\Psi = D^T(U + P) - 2I$. D can be updated by the rule:

$$D_{i,k} \leftarrow D_{i,k} \sqrt{\frac{[U + P + (D\Psi)^-]_{i,k}}{[2D + (D\Psi)^+]_{i,k}}} \quad (17)$$

D. Extensions

As mentioned, various side-information have been introduced to alleviate the data sparsity problem. The user uncertain preference information could be interpreted as an extension of side-information. Thus, we could integrate this information to further improve the performance of recommendation. For example, ConvMF utilizing user reviews as side-information has made a great success, and we could boost the performance by extending ConvMF with fuzzy information. Therefore, the objective function of the extension approach ZMF-C could be defined as:

$$\begin{aligned}
\min \sum_{i=1}^{N}\sum_{j=1}^{M} W_{i,j} \left(\left(R_{i,j} - U_i V_j^T\right)^2 + \left(S_{i,j} - P_i Q_j^T\right)^2 \right) \\
+ \sum_{j=1}^{M} \left(\lambda_V ||V_j - \text{cnn}(\Theta, X_j)||_2^2 + \lambda_Q ||Q_j||_2^2\right) + \lambda_\Theta ||\Theta||_F^2 \\
+ \sum_{i=1}^{N} \left(\lambda_U ||U_i - D_i||_2^2 + \lambda_P ||P_i - D_i||_2^2\right) \\
s.t.\ D^T D = I, D \geq 0, U \geq 0, P \geq 0 \quad (18)
\end{aligned}$$

where X_j is the textual information for item j. We can obtain the update rule for V as:

$$V_j \leftarrow (R_j W_j U + \lambda_V \text{cnn}(\Theta, X_j))(U^T W_j U + \lambda_V E_K)^{-1} \quad (19)$$

And variable Θ can be learned by the back propagation algorithm, the remaining variables can be updated as above. The complete optimization algorithm is presented in Algorithm 1. For the extension model of ConvMF, we need to input the side-information X and replace the step 10 of Algorithm 1 with Eq.(19).

E. Complexity Analysis

For each iteration of U_i and V_j, the major cost is computing the matrix inversion and matrix multiplication. We assume the time complexity of matrix inversion $(V^T W_i V)^{-1}$ is $O\left(K^3\right)$. Thus, in ZMF and ZMF-P, the running time of U_i is $O\left(K^3 + K^2 n_i\right)$ and $O(K^2 n_i)$, respectively, where n_i is the number of items rated by user i. Analogously, the running time of V_j and Q_j are both $O(K^3 + K^2 m_j)$, where m_j is the number of users who rated item j. Consequently, for N users and M items, the total time complexity is $O\left(K^3\left(N + M\right) + K^2\ell\right)$ and $O\left(K^3 M + K^2\ell\right)$ for ZMF and ZMF-P respectively, where ℓ is the overall number of observation, that is $\ell = \sum_i n_i = \sum_j m_j$.

In summary, the proposed approaches and the respective base model (BiasMF/ConvMF) belong to the same order of magnitude, i.e., it is linear with the input size. Moreover, the space complexity of the proposed approaches is also same as the respective base model, without any additional space cost.

TABLE I: Statistics of three datasets

Datasets	#Users	#Items	#Ratings	Density
ML-1M	6,040	3,544	993,482	4.641%
ML-10M	69,878	10,073	9,945,875	1.413%
AIV	29,757	15,149	135,188	0.030%

IV. Experiments

In this section, we evaluate the proposed approach from the following perspectives: 1) the effectiveness compared with other approaches; 2) the performance in sparse data situation; 3) the impact of fuzzy set parameters.

A. Datasets

We conduct experiments on three public real-world datasets: Movie-Lens 1M (ML-1M)[1], Movie-Lens 10M (ML-10M), and Amazon Instant Video (AIV)[2]. Table I summarizes the statistics on experimental datasets. These datasets all contain explicit feedback ratings on a scale of 1 to 5. For the approaches such as ConvMF, ZMF-C, we need to collect item reviews information. The item reviews of AIV are provided by itself. While Movie-Lens does not contain this information, we use the description information of item instead which could be acquired from IMDB[3]. In addition, we remove the users that have less than 3 ratings for AIV to improve the performance. In this paper, we split the original dataset into training, validation and testing sets with the 80%:10%:10% split. To demonstrate the effectiveness of the evaluation, we make sure that each part contains all users with one rating at least. Interested readers can also refer to paper [6] for other experimental details.

B. Metrics

We adopt the Root Mean Squared Error (RMSE) [25] to measure the divergences between the predicted rating and ground-truth rating. Besides, we use the Recall@m

[1]https://grouplens.org/datasets/movielens/
[2]http://jmcauley.ucsd.edu/data/amazon/
[3]http://www.imdb.com/

TABLE II: Comparison of the Used Information in Each Approach

Approaches	Rating	Text	Bias	Uncertainty
First Group				
PMF	✓	\	\	\
WNMF	✓	\	\	\
BiasMF	✓	\	✓	\
ZMF	✓	\	✓	✓
Second Group				
CTR	✓	✓	\	\
CDL	✓	✓	\	\
ConvMF	✓	✓	\	\
ZMF-C	✓	✓	✓	✓

to evaluate the performance in the top-m recommendation. They are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{|T|}\sum_{R_{i,j}\in T}(\hat{R}_{i,j}-R_{i,j})^2} \tag{20}$$

$$\text{Recall@}m = \frac{1}{N}\sum_{i}^{N}\frac{\min\left(\sum_j^M sign\left(\hat{R}_{i,j}-\mu_{inf}^i\right), m\right)}{\min\left(\sum_j^M sign\left(R_{i,j}-\mu_{inf}^i\right), m\right)} \tag{21}$$

where T is the test sets, $|T|$ is the total number of rating; sign(x) = 1 if $x > 0$, otherwise, sign(x)=0. And we note user i likes the item j if $R_{i,j} > \mu_{inf}^i$. We can obtain the final result of Recall@m by computing the average of all users.

C. Comparison Approaches

For a fair comparison, we divide the comparison approaches into two groups. In the first group, only the rating information is adopted, such as PMF [10], WNMF [26], BiasMF [1], and ZMF. In the second group, both the rating and textual information are adopted, such as CTR [4], CDL [5], ConvMF [6], and ZMF-C. The characteristics of the comparative approaches are listed in Table II.

D. Parameter Settings

The hyper parameter plays a vital role in the final performance. The PMF and ConvMF parameters are derived from the references [6]. Other model hyper parameters are obtained by the grid approach, as shown in Table 1. What's more, we reproduce the approaches of PMF, ConvMF, BiasMF, and WNMF. The experimental results of CTR and CDL are taken directly from [6]. In this paper, all variables of our approach are randomly initialized to $[0.001, 1]$. For ZMF-C, we use the CNN model, which has the same specific parameters as in ConvMF.

E. Experimental Results

In this subsection, we will show the experimental results and make some analysis. All experimental results are average values of 5 trials.

1) The effectiveness compared with other approaches: **Rating Prediction.** In Table III, we evaluate the proposed approach at different latent factors dimensions K, providing a comparison with other approaches on the three recommendation datasets. We can observe that proposed approach could achieve more excellent performance than other comparing approaches. *For the First Group,* taking the potential factor $K = 50$ for example, it can be observed that ZMF achieves the best result. Specifically, comparing with PMF, RMSE has decreased by 2.6% on ML-1M, 1.8% on ML-10M, and 7.8% on AIV. Besides, the strong baseline of BiasMF makes a more excellent performance than PMF on ML-1M and ML-10M datasets, which shows that the **bias of user and item** do affect the performance of the model. It must be said that ZMF makes a further improvement than BiasMF, i.e., RMSE has decreased by 1.6% on ML-1M, 1.5% on ML-10M, and 10.8% on AIV. The experimental results demonstrate that the user uncertain information plays an important role and brings a significant improvement. *For the Second Group,* taking the potential factor $K = 50$ for example, it can be observed that extension approach ZMF-C is also good. Comparing the results of above two group approaches, these approaches with massive side-information make a great promotion, indicating that the auxiliary information is meaningful. We also investigate the effect of the latent factors dimension K by setting K from $\{30, 50, 100\}$. It can be observed that these approaches of each group have achieved consistent results, i.e., the performance continues to improve as D increases.

Top-m recommendation. In recommender system, the users often pay more attention to the results listed in the top. Besides, the evaluation of recall denotes the purpose of recommendation. To further evaluate the effectiveness of the proposed approach, the top-m performance are shown in Fig. 3 for recommendation system. The performance is often measured by the recall rate and denoted as Recall@m. We can observe from Figure 3(a) that both ZMF and ZMF-C have obvious advantages on above three datasets. *For the First Group,* the recall of ZMF is improved by 5.6% than PMF when $m = 20$ on ML-1M. *For the Second Group,* ZMF-C also outperforms ConvMF by 8.9% when $m = 20$. Analogously, the recall of proposed approach has a consistently better than baselines (i.e.,PMF, ConvMF) on ML-10M and AIV datasets. Besides, m is set from $\{3, 5, 10, 15, 20, 25, 30, 40, 50\}$. We can observe that our proposed approach has a stable improvement. It must be said that the higher accuracy is required under the smaller value m for the top-m recommendation system. Therefore, the performance should be discussed when $m = 2$. *For the First Group,* ZMF

TABLE III: The performance in terms of RMSE on three datasets

Approaches	ML-1M			ML-10M			AIV		
	K=30	K=50	K=100	K=30	K=50	K=100	K=30	K=50	K=100
First Group									
PMF	0.9037	0.8971	0.8894	0.8311	0.8287	0.8255	1.2008	1.1889	1.1409
WNMF	0.9365	0.9296	0.9212	0.8907	0.8834	0.8793	1.2372	1.2062	1.1840
BiasMF	0.8881	0.8875	0.8743	0.8298	0.8258	0.8214	1.2206	1.2189	1.1901
ZMF	**0.8718**	**0.8703**	**0.8629**	**0.8115**	**0.8110**	**0.8064**	**1.1523**	**1.1110**	**1.0995**
Second Group									
CTR	N/A	0.8969	N/A	N/A	0.8275	N/A	N/A	1.5496	N/A
CDL	N/A	0.8879	N/A	N/A	0.8186	N/A	N/A	1.3594	N/A
ConvMF	0.8646	0.8531	0.8525	0.7978	0.7958	0.7885	1.1365	1.1337	1.1111
ZMF-C	**0.8501**	**0.8410**	**0.8402**	**0.7878**	**0.7840**	**0.7796**	**1.1000**	**1.0710**	**1.0612**

outperforms PMF by 3.9% and 3.3% on the ML-1M and ML-10M, respectively. *For the Second Group,* ZMF-C is also more excellent than ConvMF by 6.4%, 8.5% and 4.5% on three datasets respectively. The experimental results show that the proposed approach are more effective and increase the number of recalled items which user really likes.

In general, our approach achieve a promising improvement on RMSE, and recall@m compared with other approaches. And it's worth mentioning that our approach can be easy to extend to other existing approaches based on MF framework.

2) *Evaluation of Alleviating the Data Sparsity :*

In Figure 4, the effect of data sparsity is investigated for recommendation performance on ML-1M dataset. By sampling randomly from ML-1M, five additional datasets are obtained, where the proportion of the training set in each dataset is denoted as x, and $x \in \{0.2, 0.4, 0.5, 0.6, 0.8\}$. We can observe from Figure 4 that the proposed approach has a more excellent performance than baselines obviously. *For the First Group*,when $x = 0.2$, the RMSE of ZMF is about 2.1% lower than PMF, and 6.4% lower than BiasMF. *For the Second Group,* ZMF-C makes the best performance and outperforms ConvMF by 6.1% on RMSE. With the number of data increases, there are abundant reviews and rating behavioral information. In this situation, more precise item/user latent factors could be obtained, resulting in more excellent performance for recommendation. Above experimental results demonstrate that the derived information is important for alleviating the problem of data sparsity and improving the quality of recommender systems.

3) *Impact of the Fuzzy Set Parameters:*

In Table IV, the affect of fuzzy set parameters is evaluated on model performance. The most important parameters are the infimum and supremum of each user's preference. First, we sort the rating record of each user. Then, the minimum, maximum, quartile, and average value (i.e., min^i, max^i, $[\frac{1}{4}]^i or [\frac{3}{4}]^i$, $average^i$) are obtained. In this section, four pairs of parameters are utilized to

TABLE IV: The affect of the $<\mu^i_{inf},\mu^i_{sup}>$ on ML-1M and ML-10M ($K = 30$).

$<\mu^i_{inf},\mu^i_{sup}>$*	ZMF	
	ML-1M	ML-10M
$<average^i,average^i>$	0.8777	0.8201
$<min^i,max^i>$	0.8774	0.8180
$<average^i,max^i>$	0.8737	0.8150
$<[\frac{1}{4}]^i,[\frac{3}{4}]^i>$	**0.8718**	**0.8115**

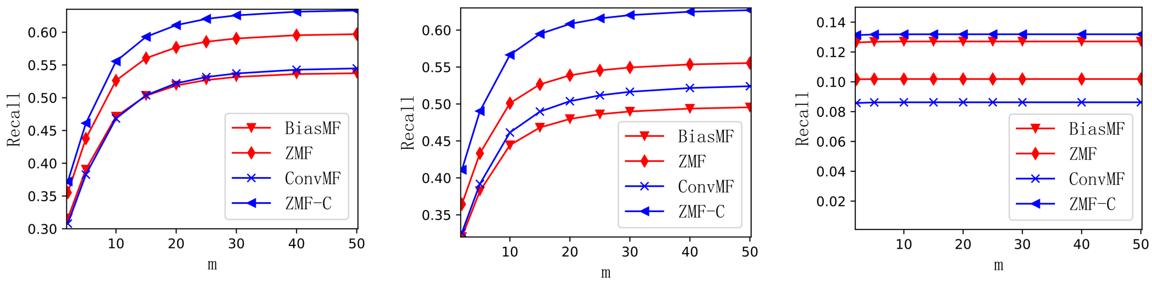
* Each user has specific value, which represents the bias of each user.

evaluate ZMF, respectively. When $\mu^i_{inf}=\mu^i_{sup}=average^i$, it indicates that the model only considers the rating bias of user and ignores the uncertainty of the middle rating. When $<\mu^i_{inf},\mu^i_{sup}>$ is $<min^i,max^i>$, it shows that all of the ratings are uncertain. Moreover, $<average^i,max^i>$ and $<[\frac{1}{4}]^i,[\frac{3}{4}]^i>$ are all the general cases, which consider the explicit preference and uncertain preference simultaneously. It can be observed from Table IV that ZMF has obvious advantages on the parameters of $<[\frac{1}{4}]^i,[\frac{3}{4}]^i>$. Specifically, the experimental results of $<average^i,average^i>$ and $<average^i,max^i>$ show that the uncertainty of middle rating is beneficial to the improvement of the model performance. Comparing with $<min^i,max^i>$ and $<average^i,max^i>$, we know that very lower ratings usually have represented the dislike of user in most cases. Above experimental results demonstrate that the user uncertain preference information is critical for recommender systems.

V. Conclusion

In this paper, we makes an initial attempt in integrating the influence of user uncertain preference information by rating bias for fuzzy recommender systems. Specifically, ZMF are proposed for formulating the *rating bias and rating uncertainty* into the matrix factorization framework. ZMF-C are proposed for extending the *uncertain preference information* into ConvMF. The uncertain preference information could reflect the user-specific

978-1-7281-6926-2/20/$31.00 ©2020 IEEE



(a) ML-1M (b) ML-10M (c) AIV

Fig. 3: The performance in terms of Recall@m on three datasets

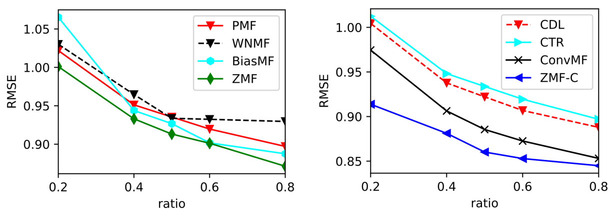


Fig. 4: The performance on various ratio of training set on ML-1M

taste accurately without introducing additional data. Experimental results show that the proposed approach produces an effective improvement compared to other approaches. In the future, the emotional polarity of user reviews will be exploited for recommendation system, simultaneously.

References

[1] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, 2009.

[2] O. Wolkenhauer, *Data engineering: fuzzy mathematics in systems theory and data analysis*. John Wiley & Sons, 2004.

[3] Z. Ma, F. Zhang, L. Yan, and J. Cheng, *Fuzzy knowledge management for the semantic web*. Springer, 2016.

[4] C. Wang and D. M. Blei, "Collaborative topic modeling for recommending scientific articles," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011, pp. 448–456.

[5] H. Wang, N. Wang, and D.-Y. Yeung, "Collaborative deep learning for recommender systems," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2015, pp. 1235–1244.

[6] D. Kim, C. Park, J. Oh, S. Lee, and H. Yu, "Convolutional matrix factorization for document context-aware recommendation," in *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 2016, pp. 233–240.

[7] E. Christakopoulou and G. Karypis, "Local item-item models for top-n recommendation," in *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 2016, pp. 67–74.

[8] D. Liang, J. Altosaar, L. Charlin, and D. M. Blei, "Factorization meets the item embedding: Regularizing matrix factorization with item co-occurrence," in *Proceedings of the 10th ACM conference on recommender systems*. ACM, 2016, pp. 59–66.

[9] D. Cao, L. Nie, X. He, X. Wei, S. Zhu, and T.-S. Chua, "Embedding factorization models for jointly recommending items and user generated lists," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2017, pp. 585–594.

[10] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," in *Advances in neural information processing systems*, 2008, pp. 1257–1264.

[11] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[12] R. Yera and L. Martinez, "Fuzzy tools in recommender systems: A survey," *Int. J. Comput. Intell. Syst*, vol. 10, no. 1, pp. 776–803, 2017.

[13] R. M. Rodriguez, L. Martinez, and F. Herrera, "Hesitant fuzzy linguistic term sets for decision making," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 1, pp. 109–119, 2012.

[14] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—i," *Information sciences*, vol. 8, no. 3, pp. 199–249, 1975.

[15] A. Zenebe and A. F. Norcio, "Representation, similarity measures and aggregation methods using fuzzy sets for content-based recommender systems," *Fuzzy sets and systems*, vol. 160, no. 1, pp. 76–94, 2009.

[16] R. R. Yager, "Fuzzy logic methods in recommender systems," *Fuzzy Sets and Systems*, vol. 136, no. 2, pp. 133–149, 2003.

[17] L.-C. Cheng and H.-A. Wang, "A novel fuzzy recommendation system integrated the experts' opinion," in *Fuzzy Systems (FUZZ), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2060–2065.

[18] L. A. Zadeh, "Fuzzy sets," in *Fuzzy Sets, Fuzzy Logic, And Fuzzy Systems: Selected Papers by Lotfi A Zadeh*. World Scientific, 1996, pp. 394–432.

[19] H. Liu and H.-A. Jacobsen, "Modeling uncertainties in publish/subscribe systems," in *Data Engineering, 2004. Proceedings. 20th International Conference on*. IEEE, 2004, pp. 510–521.

[20] F. Herrera and L. Martínez, "A 2-tuple fuzzy linguistic representation model for computing with words," *IEEE Transactions on fuzzy systems*, vol. 8, no. 6, pp. 746–752, 2000.

[21] Z. Ma, *A Literature Overview of Fuzzy Database Modeling*. IGI Global, 2009.

[22] C. H. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 1, pp. 45–55, 2010.

[23] C. Ding, T. Li, W. Peng, and H. Park, "Orthogonal nonnegative matrix t-factorizations for clustering," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2006, pp. 126–135.

[24] P. Richtárik and M. Takáč, "Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function," *Mathematical Programming*, vol. 144, no. 1-2, pp. 1–38, 2014.

[25] G. Guo, J. Zhang, and N. Yorke-Smith, "A novel recommendation model regularized with user trust and item ratings," *ieee transactions on knowledge and data engineering*, vol. 28, no. 7, pp. 1607–1620, 2016.

[26] F. Zhang, N. J. Yuan, D. Lian, X. Xie, and W.-Y. Ma, "Collaborative knowledge base embedding for recommender systems," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2016, pp. 353–362.