

A Fuzzy Set Based Approach for Rating Bias

Mingming Li,^{1,2,3} Jiao Dai,^{1*} Fuqing Zhu,¹ Liangjun Zang,¹ Songlin Hu,^{1,3} Jizhong Han¹

¹Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

²School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

³University of Chinese Academy of Sciences, Beijing, China

{limingming,daijiao,zhufuqing,zangliangjun,husonglin,hanjizhong}@iie.ac.cn

Abstract

In recommender systems, the user uncertain preference results in unexpected ratings. This paper makes an initial attempt in integrating the influence of user uncertain degree into the matrix factorization framework. Specifically, a fuzzy set of *like* for each user is defined, and the membership function is utilized to measure the degree of an item belonging to the fuzzy set. Furthermore, to enhance the computational effect on sparse matrix, the uncertain preference is formulated as a side-information for fusion. Experimental results on three real-world datasets show that the proposed approach produces stable improvements compared with others.

Introduction

Matrix Factorization is one of the most famous and successful recommendation approaches, which learns the discriminative latent factors for users and items by factorizing user-item interaction matrix. However, there still exists a drawback: The rating information is uncertain and imprecise. Different users have various evaluation criteria, and different ratings also have different degrees of preferences.

To handle the uncertain and imprecise information of rating, the user/item bias vector is utilized to adjust original ratings, such as BiasMF (Koren, Bell, and Volinsky 2009). Nevertheless, the impact of uncertain ratings is neglected. Specifically, as shown in Figure 1, items with very low/high ratings explicitly indicate that user really dislikes/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 differentiate members belong to the set of *like* (e.g., rating>4) from non-members.

*Corresponding authors

Copyright © 2019, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

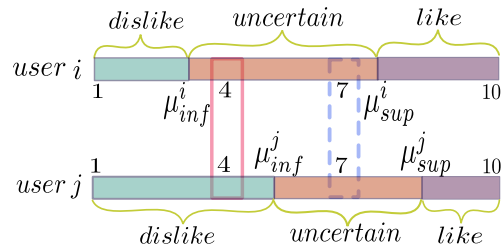


Figure 1: Illustration of rating bias.

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* by fuzzy set theory (Ma et al. 2016).

Given the above considerations, we explore the user uncertain preference information using fuzzy set and make them as the side-information. Then an approach of sharing user latent factors with fuZzy Matrix Factorization (ZMF) is proposed. Experiments show that the proposed approach consistently achieves 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.
- Fuzzy matrix factorization based approach is proposed to integrate user uncertain preference and rating information, which could overcome the uncertain ratings and data sparsity problems.
- On three public explicit feedback datasets, extensive experimental results demonstrate that the proposed approach produces competitive performances.

Proposed Approach

Suppose there are N users, M items, and a user-item rating matrix $R \in \mathcal{R}^{N \times M}$. Let $R_{i,j}$ represent 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 $K \ll \min(M, N)$.

User Uncertain Preference Matrix Construction

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 $\tilde{F}_i, \tilde{F} = \{\tilde{F}_i | i = 1, \dots, N\}$. We use the S-type function to measure the membership degrees of \tilde{F}_i :

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

where $a > 0, b < 0, \xi_i > 0$, and μ_{inf}^i and μ_{sup}^i is the infimum and supremum of user i respectively. To simplify our model, we set $a = \mu_{sup}^i - \mu_{inf}^i + 1, b = -3$, and $\xi_i = a(\mu_{sup}^i - \mu_{inf}^i)^b$.

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

$$S_{i,j} = \psi_{\tilde{F}_i}(R_{i,j}) \quad (2)$$

fuZzy Matrix Factorization

To integrate the user preference information into MF framework, we propose a fuzzy set based approach, which jointly decomposes the user-item interaction matrix and the user uncertain preference matrix with sharing user latent vector. This framework is effective, and has achieved great success in many models (Zhang et al. 2014; Cao et al. 2017).

The objective function of ZMF could be expressed as follows:

$$\min_{U,V,Q} \mathcal{L} = \|W \odot (R - UV^T)\|_F^2 + \|W \odot (S - UQ^T)\|_F^2 + \lambda_U \|U\|_F^2 + \lambda_V \|V\|_F^2 + \lambda_Q \|Q\|_F^2 \quad (3)$$

We could use the block-coordinate descent algorithm to find optimal solutions for U, V, Q .

Experiments

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. 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 the users with one rating at least. We adopt the Root Mean Squared Error (RMSE) (Guo, Zhang, and Yorke-Smith 2016) and Mean Absolute Error (MAE) to measure the divergences between the predicted rating and ground-truth rating. The parameters of

¹<https://grouplens.org/datasets/movielens/>

²<http://jmcauley.ucsd.edu/data/amazon/>

Table 1: Comparison Results.

Datasets		PMF	WNMF	BiasMF	ZMF
ML-1M	RMSE	0.9037	0.9365	0.8881	0.8752
	MAE	0.7031	0.7424	0.6929	0.6881
ML-10M	RMSE	0.8311	0.8907	0.8298	0.8156
	MAE	0.6395	0.6903	0.6332	0.6240
AIV	RMSE	1.2008	1.2372	1.2206	1.1882
	MAE	0.8608	0.8655	0.8492	0.8353

PMF are derived from the references (Kim et al. 2016), and other models parameters are obtained by the grid approach. All variables of our approaches are randomly initialized to $[0.001, 1]$.

As shown in Table 1, ZMF achieves an excellent performance compared with other approaches. More importantly, our approach makes a further improvement than BiasMF, which could explain that the user uncertain information plays an important role and brings a significant improvement.

Conclusion

In this paper, we explore the user uncertain preference information from original rating data to handle the fuzzy rating problem. The uncertain preference information could reflect the user-specific taste accurately. Extensive experiments conducted on three real-world explicit feedback datasets demonstrate that the proposed approach produces significant improvements compared with others.

References

- Cao, D.; Nie, L.; He, X.; Wei, X.; Zhu, S.; and Chua, T.-S. 2017. 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*, 585–594. ACM.
- Guo, G.; Zhang, J.; and Yorke-Smith, N. 2016. A novel recommendation model regularized with user trust and item ratings. *IEEE transactions on knowledge and data engineering* 28(7):1607–1620.
- Kim, D.; Park, C.; Oh, J.; Lee, S.; and Yu, H. 2016. Convolutional matrix factorization for document context-aware recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems*, 233–240. ACM.
- Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer* 42(8).
- Ma, Z.; Zhang, F.; Yan, L.; and Cheng, J. 2016. *Fuzzy knowledge management for the semantic web*. Springer.
- Zhang, Y.; Lai, G.; Zhang, M.; Zhang, Y.; Liu, Y.; and Ma, S. 2014. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, 83–92. ACM.