

k-CoFi: Modeling *k*-Granularity Preference Context in Collaborative Filtering

Yunfeng Huang¹, Zixiang Chen¹, Lin Li¹, WeiKe Pan^{1(✉)}, Zhiguang Shan²,
and Zhong Ming^{1(✉)}

¹ College of Computer Science and Software Engineering,
Shenzhen University, Shenzhen, China

{huangyunfeng2017, chenzixiang2016, lilin20171}@email.szu.edu.cn,
{panweike, mingz}@szu.edu.cn

² Information Research Department, State Information Center, Beijing, China
shanzhiguang@263.com

Abstract. Collaborative filtering (CF) is a highly applicable technology for predicting a user's rating to a certain item. Recently, some works have gradually switched from modeling users' rating behaviors alone to modeling both users' behaviors and preference context beneath rating behaviors such as the set of other items rated by user u . In this paper, we go one step beyond and propose a novel perspective, i.e., k -granularity preference context, which is able to absorb existing preference context as special cases. Based on this new perspective, we further develop a novel and a generic recommendation method called k -CoFi that models k -granularity preference context in collaborative filtering in a principled way. Empirically, we study the effectiveness of factorization with coarse granularity, fine granularity and smooth granularity, and their complementarity, by applying k -CoFi to three real-world datasets. We also obtain some interesting and promising results and useful guidance for practitioners from the experiments.

Keywords: k -granularity preference context · Matrix factorization
Collaborative filtering

1 Introduction

Collaborative filtering (CF) or (user, item) rating prediction is one of the most well-known and well-studied technology in recommender systems largely because of its high applicability to different applications and domains. Users' preference modeling is one of the most fundamental issues in developing advanced CF methods such as factorization machine [4] and deep learning [1].

Recently, some works have switched to modeling both users' preferences and their preference context contained in the rating data, instead of modeling users' preferences alone. For example, in SVD++ [2], in order to predict the rating of a (user, item) pair (u, i) , besides the latent representation of user u and item i ,

the latent representation of other items rated by user *u* are also included as the preference context. The justification is that two users with similar sets of rated items are likely to have similar preference patterns. As another instance, in MF-MPC (matrix factorization with multiclass preference context) [3], the preference context of rated items in SVD++ [2] is further refined according to the rating values of the user to other items, which is able to capture more accurate preference context.

However, these works are studied independently and separately, and have not been studied in a generic and unified framework. For this reason, their relationship and potential of combination also have not been fully exploited. In this paper, we propose a novel perspective, i.e., *k*-granularity preference context, which provides a unified view and absorbs different preference context in existing works as special cases. Based on this new perspective, we develop a novel algorithm that models *k*-granularity preference context in collaborative filtering (*k*-CoFi). Furthermore, we study the complementarity of different *k*-granularity preference context in factorization-based algorithms.

In order to study the effectiveness and complementarity of different *k*-granularity preference context. We conduct extensive empirical studies on three real-world datasets with several state-of-the-art methods. The results clearly show the effectiveness of fine granularity and smooth granularity, and their complementarity in rating prediction.

We summarize our main contributions as follows: (i) we propose a novel perspective for modeling users’ preference context, i.e., *k*-granularity preference context; (ii) we develop a novel and generic recommendation algorithm that models *k*-granularity preference context in collaborative filtering, i.e., *k*-CoFi; and (iii) we study the effectiveness and complementarity of different preference context and have some interesting observations.

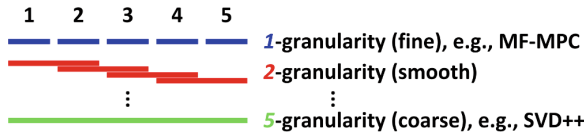


Fig. 1. Illustration of different granularity of 5-star numerical ratings in collaborative filtering.

2 *k*-Granularity Preference Context

In this section, we describe a novel perspective, i.e., *k*-granularity preference context, for modeling users’ preference context in collaborative filtering. Firstly, we formally define *k*-granularity preference context in collaborative filtering with categorical rating values. Secondly, we review two state-of-the-art collaborative filtering methods, i.e., SVD++ [2] and MF-MPC [3], from the perspective of virtual user profiles based on 5-granularity preference context and 1-granularity preference context, respectively.

We list some commonly used notations in Table 1.

Table 1. Some notations and explanations.

n	user number
m	item number
$u \in \{1, 2, \dots, n\}$	user ID
$i \in \{1, 2, \dots, m\}$	item ID
$\mathbb{V} = \{v\}$	a set of rating values
$r_{ui} \in \mathbb{V}$	rating assigned by user u to item i
$\mathcal{R} = \{(u, i, r_{ui})\}$	training data of rating records
\mathcal{I}_u	a set of items w.r.t. u
\mathcal{I}_u^v	a set of items w.r.t. u and v
$\mathcal{K} = \{k\}$	a set of k -granularity
$\mathcal{S}_k = \{g\}$	a set of rating groups w.r.t. k
\mathcal{I}_u^{kg}	a set of items w.r.t. u , k and g
d	size of latent vector
$\mu \in \mathbb{R}$	global average rating value
$b_u \in \mathbb{R}$	user bias of user u
$b_i \in \mathbb{R}$	item bias of item i
$U_u. \in \mathbb{R}^{1 \times d}$	latent vector of user u
$V_i. \in \mathbb{R}^{1 \times d}$	latent vector of item i
$W_{i'}^{kg} \in \mathbb{R}^{1 \times d}$	latent vector of item i' w.r.t. k and g
\hat{r}_{ui}	predicted rating of user u to item i
T	iteration number in the algorithm
α	weight on regularization terms
λ	weight for combining two k -CoFi

2.1 Definition of Preference Context

In collaborative filtering, we usually have a set of multiclass rating values such as $\{1, 2, 3, 4, 5\}$ in MovieLens 100K and MovieLens 1M. In order to model a rating of a user u to an item i , we can represent it as follows,

$$P(r_{ui}|(u, i, \mathcal{C})), \quad (1)$$

where \mathcal{C} denotes the preference context from the rating data itself rather than some auxiliary temporal or spatial context. In particular, the context can be (i) $\mathcal{C} = \emptyset$ in PMF [6]; (ii) $\mathcal{C} = \mathcal{I}_u \setminus \{i\}$ denoting the rated items (excluding item i itself) by user u in SVD++ [2]; and (iii) $\mathcal{C} = \mathcal{I}_u^v \setminus \{i\}$, $v \in \mathbb{V}$ denoting the rated items with different rating values $v \in \mathbb{V}$ (excluding item i itself) by user u in MF-MPC [3].

We may view the preference context from a new perspective, i.e., from coarse granularity to fine granularity. For example, (i) $\mathcal{C} = \emptyset$ represents none context, i.e., 0-granularity; (ii) $\mathcal{C} = \mathcal{I}_u \setminus \{i\}$ represents 5-granularity meaning without distinguishing the rating values; and (iii) $\mathcal{C} = \mathcal{I}_u^v \setminus \{i\}, v \in \mathbb{V}$ represents 1-granularity meaning treating each rating value separately. We illustrate this new perspective in Fig. 1.

Following this line of discussion, we propose a novel and generic preference context, i.e., *k*-granularity preference context, in which *k* consecutive rating values are treated as one group of ratings without differences. For example, the set of rating groups in 5-granularity and 1-granularity are $\mathcal{S}_5 = \{\{1, 2, 3, 4, 5\}\}$ and $\mathcal{S}_1 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}$, respectively.

We can see that \mathcal{S}_5 and \mathcal{S}_1 are two extreme cases, one for coarse granularity and the other for fine granularity. Naturally, we may consider the preference context between these two cases, e.g., 2-granularity preference context with $\mathcal{S}_2 = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 5\}\}$. We can take the 2-granularity preference context as a smooth preference context, where we do not distinguish two consecutive rating values. This makes sense because sometimes the difference between two consecutive rating values may not be so significant for a user, especially when a user *u* is affected by his or her own mood in different situations.

In the following two subsections, we will show how the *k*-granularity preference context is modeled for rating prediction in two representative works, i.e., SVD++ [2] and MF-MPC [3]. In particular, we will represent the modeling technique by a virtual user profile in each case.

2.2 Virtual User Profile in SVD++

In order to capture the preference context of rated items by the current end user *u*, a virtual user profile is introduced in SVD++ [2],

$$\begin{aligned}
 \tilde{U}_u^{\text{SVD}++} &= \frac{1}{\sqrt{|\mathcal{I}_u \setminus \{i\}|}} \sum_{i' \in \mathcal{I}_u \setminus \{i\}} W_{i'} \\
 &= \sum_{g \in \mathcal{S}_5} \frac{1}{\sqrt{|\mathcal{I}_u^{5g} \setminus \{i\}|}} \sum_{i' \in \mathcal{I}_u^{5g} \setminus \{i\}} W_{i'}^{5g} \\
 &= \sum_{k \in \{5\}} \sum_{g \in \mathcal{S}_k} \frac{1}{\sqrt{|\mathcal{I}_u^{kg} \setminus \{i\}|}} \sum_{i' \in \mathcal{I}_u^{kg} \setminus \{i\}} W_{i'}^{kg} \tag{2}
 \end{aligned}$$

It is clear that the virtual user profile $\tilde{U}_u^{\text{SVD}++}$ in Eq. (2) is based on the 5-granularity preference context.

2.3 Virtual User Profile in MF-MPC

Similarly, in order to model the fine-granularity preference context of rated items with different values $v \in \mathbb{V}$, a sophisticated virtual user profile of user *u* is used in MF-MPC [3],

$$\begin{aligned}
\tilde{U}_{u.}^{\text{MF-MPC}} &= \sum_{v \in \mathbb{V}} \frac{1}{\sqrt{|\mathcal{I}_u^v \setminus \{i\}|}} \sum_{i' \in \mathcal{I}_u^v \setminus \{i\}} W_{i'}^v. \\
&= \sum_{g \in \mathcal{S}_1} \frac{1}{\sqrt{|\mathcal{I}_u^{1g} \setminus \{i\}|}} \sum_{i' \in \mathcal{I}_u^{1g} \setminus \{i\}} W_{i'}^{1g}. \\
&= \sum_{k \in \{1\}} \sum_{g \in \mathcal{S}_k} \frac{1}{\sqrt{|\mathcal{I}_u^{kg} \setminus \{i\}|}} \sum_{i' \in \mathcal{I}_u^{kg} \setminus \{i\}} W_{i'}^{kg}. \tag{3}
\end{aligned}$$

We can see that the virtual user profile $\tilde{U}_{u.}^{\text{MF-MPC}}$ precisely captures the information encoded in the 1-granularity preference context.

3 k-CoFi

3.1 Problem Definition

In collaborative filtering, the main task is to learn users' preferences from (user, item, rating) triples in the training data, and then predict the rating of each (user, item) pair in the test data.

3.2 Preference Context in k-CoFi

With the k -granularity preference context, we have the generic latent representation of a virtual user profile as follows,

$$\tilde{U}_{u.}^{k\text{-gran}} = \sum_{k \in \mathcal{K}} \sum_{g \in \mathcal{S}_k} \frac{1}{\sqrt{|\mathcal{I}_u^{kg} \setminus \{i\}|}} \sum_{i' \in \mathcal{I}_u^{kg} \setminus \{i\}} W_{i'}^{kg}, \tag{4}$$

where \mathcal{K} denotes a set of k -granularity and \mathcal{S}_k is a set of rating groups w.r.t. k .

We can see that $\tilde{U}_{u.}^{k\text{-gran}}$ in Eq. (4) reduces to $\tilde{U}_{u.}^{\text{SVD}++}$ in Eq. (2) when $\mathcal{K} = \{5\}$, and reduces to $\tilde{U}_{u.}^{\text{MF-MPC}}$ in Eq. (3) when $\mathcal{K} = \{1\}$, which shows that our proposed virtual user profile is a very generic one.

3.3 Prediction Rule

With the virtual user profile, we may define the predicted rating of user u to item i in a similar way to that of SVD++ [2] and MF-MPC [3],

$$\hat{r}_{ui} = (U_{u.} + \tilde{U}_{u.}^{k\text{-gran}})V_i^T + b_i + b_u + \mu, \tag{5}$$

where $U_{u.} + \tilde{U}_{u.}^{k\text{-gran}} \in \mathbb{R}^{1 \times d}$ denotes the overall profile of user u , $V_i. \in \mathbb{R}^{1 \times d}$ is the item-specific latent feature vector, and $b_i \in \mathbb{R}$, $b_u \in \mathbb{R}$, $\mu \in \mathbb{R}$ are item bias, user bias and global average rating, respectively.

3.4 Objective Function

We embed the prediction rule in a commonly used objective function with square loss for rating prediction,

$$\min_{\Theta} \sum_{u=1}^n \sum_{i=1}^m y_{ui} \left[\frac{1}{2} (r_{ui} - \hat{r}_{ui})^2 + \text{Reg}(u, i) \right], \tag{6}$$

where $\text{Reg}(u, i) = \frac{\alpha}{2} \|U_u\|^2 + \frac{\alpha}{2} \|V_i\|^2 + \frac{\alpha}{2} \|b_u\|^2 + \frac{\alpha}{2} \|b_i\|^2 + \sum_{k \in \mathcal{K}} \sum_{g \in \mathcal{S}_k} \sum_{i' \in \mathcal{I}_u^{kg} \setminus \{i\}} \frac{\alpha}{2} \|W_{i'}^{kg}\|^2$ is the regularization term used to avoid overfitting. And $\Theta = \{U_u, V_i, b_u, b_i, W_{i'}^{kg}\}, u = 1, 2, \dots, n, i = 1, 2, \dots, m, i' \in \mathcal{I}_u^{kg} \setminus \{i\}, k \in \mathcal{K}, g \in \mathcal{S}_k$ are model parameters to be learned.

3.5 Gradients

Denoting $f_{ui} = \frac{1}{2} (r_{ui} - \hat{r}_{ui})^2 + \text{Reg}(u, i)$ as the tentative objective function for the rating triple (u, i, r_{ui}) , we have the gradients of the model parameters,

$$\nabla U_u = -e_{ui} V_i + \alpha U_u, \tag{7}$$

$$\nabla V_i = -e_{ui} (U_u + \tilde{U}_u^{k\text{-gran}}) + \alpha V_i, \tag{8}$$

$$\nabla b_i = -e_{ui} + \alpha b_i, \tag{9}$$

$$\nabla b_u = -e_{ui} + \alpha b_u, \tag{10}$$

$$\nabla \mu = -e_{ui}, \tag{11}$$

$$\nabla W_{i'}^{kg} = -\frac{e_{ui} V_i}{\sqrt{|\mathcal{I}_u^{kg} \setminus \{i\}|}} + \alpha W_{i'}^{kg}, \tag{12}$$

where $i' \in \mathcal{I}_u^{kg} \setminus \{i\}, k \in \mathcal{K}, g \in \mathcal{S}_k$, and $e_{ui} = r_{ui} - \hat{r}_{ui}$. Notice that the gradients are the same with that of MF-MPC [3] except the virtual user profile $\tilde{U}_u^{k\text{-gran}}$ in Eq. (8), and $W_{i'}^{kg}$ in Eq. (12) with different granularity *k* and rating group *g*.

3.6 Update Rule

Finally, we have the update rules,

$$\theta = \theta - \gamma \nabla \theta \tag{13}$$

where γ is the learning rate, and θ can be U_u, V_i, b_u, b_i, μ and $W_{i'}^{kg}, i' \in \mathcal{I}_u^{kg} \setminus \{i\}, k \in \mathcal{K}, g \in \mathcal{S}_k$.

3.7 Algorithm

We formally depict the learning procedure in an algorithm shown in Fig. 2, which mainly contains two loops. In the outer loop, we go through the whole set of training rating records *T* times, and gradually decrease the learning rate via

$\gamma = \gamma \times 0.9$ when we have learned more about the model parameters. In each of the $|\mathcal{R}|$ iterations in the inner loop, we randomly take a rating record (u, i, r_{ui}) from \mathcal{R} for calculating the corresponding gradients and updating the model parameters.

The learning procedure in Fig. 2 is adopted in typical SGD-based implementations of matrix factorization. The increased time complexity during the learning procedure, in comparison with that of PMF [6], is mainly from the k -granularity preference context, because we have to take the rated items as preference context into account. However, during the test period of rating prediction, the time complexity is similar to that of PMF [6] since we can calculate the virtual user profile of each user in advance.

```

1: Initialize the model parameters
2: for  $t = 1, \dots, T$  do
3:   for  $t_2 = 1, \dots, |\mathcal{R}|$  do
4:     Randomly pick up a  $(u, i, r_{ui})$  triple from  $\mathcal{R}$ 
5:     Calculate the gradients via Eq.(7-12)
6:     Update the parameters via Eq.(13)
7:   end for
8:   Decrease the learning rate  $\gamma \leftarrow \gamma \times 0.9$ 
9: end for

```

Fig. 2. The algorithm of k -CoFi.

4 Experimental Results

4.1 Datasets and Evaluation Metrics

In our empirical studies, we use three real-world datasets, i.e., MovieLens 100K (ML100K), MovieLens 1M (ML1M) and MovieLens 10M (ML10M), which have been used in a closely related work MF-MPC [3]. In particular, ML100K and ML1M contain about 100,000 records and 1,000,000 records respectively with rating values of $\mathbb{V} = \{1, 2, 3, 4, 5\}$, and ML10M contains about 10,000,000 records with rating values of $\mathbb{V} = \{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$. Similar to the definition of different k -granularity of ML100K and ML1M in Sect. 2, we have 5-granularity, 1-granularity and 2-granularity of ML10M as follows: $\mathcal{S}_1 = \{ \{0.5\}, \{1\}, \{1.5\}, \{2\}, \{2.5\}, \{3\}, \{3.5\}, \{4\}, \{4.5\}, \{5\} \}$, $\mathcal{S}_2 = \{ \{0.5, 1, 1.5, 2\}, \{1.5, 2, 2.5, 3\}, \{2.5, 3, 3.5, 4\}, \{3.5, 4, 4.5, 5\} \}$, and $\mathcal{S}_5 = \{ \{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\} \}$. Notice that there are $n = 943$ users and $m = 1,682$ items in ML100K, $n = 6,040$ users and $m = 3,952$ items in ML1M, and $n = 71,567$ users and $m = 10,681$ items in ML10M.

For each dataset, we randomly take 80% rating records as training data and the remaining 20% rating records as test data. We repeat this procedure for five times and have five copies of training data and test data for each dataset.

As for performance evaluation, we adopt two commonly used evaluation metrics, i.e., mean absolute error (MAE) and root mean square error (RMSE).

4.2 Baselines and Parameter Settings

In our empirical studies, our main purpose is to study the effectiveness of different *k*-granularity preference context and their complementarity. For this reason, we include the following methods in our experiments:

- $\mathcal{K} = \emptyset$ denoting matrix factorization without preference context, i.e., PMF [6];
- $\mathcal{K} = \{5\}$ denoting matrix factorization with 5-granularity coarse preference context, i.e., SVD++ [2];
- $\mathcal{K} = \{1\}$ denoting matrix factorization with 1-granularity fine preference context, i.e., MF-MPC [3]; and
- $\mathcal{K} = \{2\}$ denoting matrix factorization with 2-granularity smooth preference context.

In our preliminary empirical studies, we find that factorization with $\mathcal{K} = \{1\}$ and $\mathcal{K} = \{2\}$ perform the best and are much better than that with $\mathcal{K} = \{5\}$. Thus, we further study their complementarity via two approaches: (i) $\mathcal{K} = \{1, 2\}$, and (ii) combine the predicted ratings of $\mathcal{K} = \{1\}$ and $\mathcal{K} = \{2\}$ via a weighted combination, i.e., $\lambda \hat{r}_{ui}^{\{1\}} + (1 - \lambda) \hat{r}_{ui}^{\{2\}}$.

For parameter configuration in *k*-CoFi, we follow MF-MPC [3]. Specifically, we fix the number of latent dimensions $d = 20$, iteration number $T = 50$, and search the best tradeoff parameter $\alpha \in \{0.001, 0.01, 0.1\}$.

4.3 Results

We report the main results in Table 2, from which we can have the following observations:

- Matrix factorization with preference context of $\mathcal{K} = \{5\}$, $\mathcal{K} = \{1\}$, $\mathcal{K} = \{2\}$ performs better than that without preference context, i.e., $\mathcal{K} = \emptyset$, which clearly shows the effectiveness of modeling the preference context in rating prediction;
- Matrix factorization with preference context of $\mathcal{K} = \{1\}$, $\mathcal{K} = \{2\}$ is much better than that with $\mathcal{K} = \{5\}$, which shows the effectiveness of fine and smooth preference context in comparison with the coarse one;
- The performance of matrix factorization with preference context of $\mathcal{K} = \{2\}$ is close to the very strong baseline with $\mathcal{K} = \{1\}$ (i.e., MF-MPC [3]) across three datasets, which shows the effectiveness of the smooth preference context;
- As for the combination of $\mathcal{K} = \{1\}$ and $\mathcal{K} = \{2\}$, we find that the performance can be further improved in most cases, which showcases the complementarity of the two best performing methods associated with the fine granularity and smooth granularity preference context.

Table 2. Recommendation performance of k -CoFi with different values of k . The significantly best results are marked in bold ($p < 0.01$). The values of the tradeoff parameter α or the combination weight λ are also included for reproducibility.

Data	Method	α, λ	MAE	RMSE
ML100K	$k \in \mathcal{K} = \emptyset$, i.e., PMF	$\alpha = 0.1$	0.7475 ± 0.0031	0.9447 ± 0.0031
	$k \in \mathcal{K} = \{5\}$, i.e., SVD++	$\alpha = 0.001$	0.7268 ± 0.0032	0.9253 ± 0.0034
	$k \in \mathcal{K} = \{1\}$, i.e., MF-MPC	$\alpha = 0.001$	0.7087 ± 0.0030	0.9086 ± 0.0031
	$k \in \mathcal{K} = \{2\}$	$\alpha = 0.001$	0.7090 ± 0.0031	0.9073 ± 0.0029
	$k \in \mathcal{K} = \{1, 2\}$	$\alpha = 0.01$	0.7084 ± 0.0037	0.9072 ± 0.0033
	$k \in \mathcal{K} = \{1\}, k \in \mathcal{K} = \{2\}$	$\lambda = 0.5$	0.7068 ± 0.0029	0.9051 ± 0.0027
ML1M	$k \in \mathcal{K} = \emptyset$, i.e., PMF	$\alpha = 0.001$	0.6960 ± 0.0013	0.8840 ± 0.0017
	$k \in \mathcal{K} = \{5\}$, i.e., SVD++	$\alpha = 0.001$	0.6656 ± 0.0016	0.8514 ± 0.0019
	$k \in \mathcal{K} = \{1\}$, i.e., MF-MPC	$\alpha = 0.01$	0.6599 ± 0.0014	0.8441 ± 0.0019
	$k \in \mathcal{K} = \{2\}$	$\alpha = 0.001$	0.6586 ± 0.0009	0.8466 ± 0.0015
	$k \in \mathcal{K} = \{1, 2\}$	$\alpha = 0.01$	0.6578 ± 0.0014	0.8426 ± 0.0019
	$k \in \mathcal{K} = \{1\}, k \in \mathcal{K} = \{2\}$	$\lambda = 0.5$	0.6557 ± 0.0011	0.8404 ± 0.0016
ML10M	$k \in \mathcal{K} = \emptyset$, i.e., PMF	$\alpha = 0.01$	0.6069 ± 0.0005	0.7913 ± 0.0007
	$k \in \mathcal{K} = \{5\}$, i.e., SVD++	$\alpha = 0.01$	0.6030 ± 0.0004	0.7869 ± 0.0006
	$k \in \mathcal{K} = \{1\}$, i.e., MF-MPC	$\alpha = 0.01$	0.5945 ± 0.0003	0.7779 ± 0.0004
	$k \in \mathcal{K} = \{2\}$	$\alpha = 0.01$	0.5964 ± 0.0006	0.7800 ± 0.0007
	$k \in \mathcal{K} = \{1, 2\}$	$\alpha = 0.01$	0.5948 ± 0.0004	0.7783 ± 0.0006
	$k \in \mathcal{K} = \{1\}, k \in \mathcal{K} = \{2\}$	$\lambda = 0.5$	0.5944 ± 0.0004	0.7776 ± 0.0005

In order to further study the complementarity of matrix factorization with $\mathcal{K} = \{1\}$ and $\mathcal{K} = \{2\}$, we change the value of the combination weight $\lambda \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. Notice that the hybrid method reduces to MF-MPC [3] when $\lambda = 1$ and to that with $\mathcal{K} = \{2\}$ when $\lambda = 0$. We report the results of RMSE in Fig. 3. Notice that the results of MAE are similar, which are thus not included in the paper. We can see: (i) the hybrid method performs the best with λ around 0.5, which shows the complementarity of the two factorization

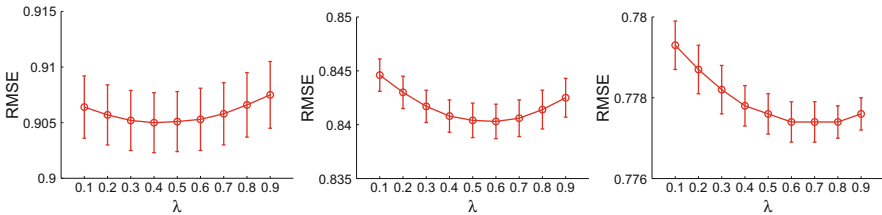


Fig. 3. Recommendation performance of combining 1-granularity and 2-granularity preference context with different values of $\lambda \in \{0.1, 0.2, \dots, 0.9\}$.

methods with different *k*-granularity again; and (ii) the trend also suggests an appropriate choice of the value of λ in practice when deploying the proposed method *k*-CoFi, i.e., practitioners may safely choose to configure it as $\lambda = 0.5$ or so.

5 Related Work

Collaborative filtering or preference prediction in a (user, item) numerical rating matrix has been studied for more than two decades. In this long journey, some seminal algorithms have been developed such as neighborhood-based methods [5] in 1990s, factorization-based methods [6] in 2000s, and deep learning based methods [1] very recently.

In neighborhood-based methods, we usually calculate the (user, user) or (item, item) similarity and construct the neighborhood for each user or item firstly, and then predict the rating for each (user, item) pair by aggregating the preferences of the users or items in the corresponding neighborhood. For either user-oriented methods or item-oriented methods, the similarity measurement and the size of neighborhood are probably two most important factors, which are usually dependent on the data properties such as the number of users, the number of items, and the density of the (user, item) rating matrix.

In factorization-based methods, we turn to predict the missing ratings in a (user, item) rating matrix via reconstructing some decomposed or factorized latent matrices. Our *k*-CoFi also belongs to this family but with a very general prediction rule based on the *k*-granularity preference context.

In deep learning-based methods, more than one copies of the factored latent matrices are learned in multiple layers of projection or embedding. In particular, some non-linear activation functions and several layers of projection make it very powerful in recommender systems beyond computer vision, speech recognition and other areas.

In this paper, we study *k*-granularity preference context, which is vertical to the above three categories of representative methods. In particular, our *k*-granularity may also be applied to improve neighborhood-based methods and deep learning based methods via estimating the similarity more accurately when constructing a better neighborhood or learning better latent representations in a neural network.

6 Conclusions and Future Work

In this paper, we study a classical rating prediction problem in collaborative filtering from a novel perspective. Specifically, we propose a new *k*-granularity preference context, which absorbs existing fine and coarse preference context as special cases. We further develop a novel and generic recommendation algorithm, i.e., *k*-CoFi, with the proposed *k*-granularity preference context. Finally, we conduct extensive empirical studies, and verify the effectiveness of different

k -granularity and their complementarity, and have some interesting and useful observations.

For future works, we are interested in generalizing the k -granularity preference context with structure information to other recommendation paradigms such as neighborhood- and deep learning based collaborative filtering.

Acknowledgements. We thank the support of National Natural Science Foundation of China No. 61502307, No. 61672358 and U1636202, and Natural Science Foundation of Guangdong Province No. 2014A030310268 and No. 2016A030313038.

References

1. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., Chua, T.-S.: Neural collaborative filtering. In: Proceedings of the 26th International Conference on World Wide Web, WWW 2017, pp. 173–182 (2017)
2. Koren, Y.: 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 2008, pp. 426–434 (2008)
3. Pan, W., Ming, Z.: Collaborative recommendation with multiclass preference context. *IEEE Intell. Syst.* **32**(2), 45–51 (2017)
4. Rendle, S.: Factorization machines with libfm. *ACM Trans. Intell. Syst. Technol.* **3**(3), 57:1–57:22 (2012)
5. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: Grouplens: an open architecture for collaborative filtering of netnews. In: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, CSCW 1994, pp. 175–186 (1994)
6. Salakhutdinov, R., Mnih, A.: Probabilistic matrix factorization. In: Annual Conference on Neural Information Processing Systems, NIPS 2008, pp. 1257–1264 (2008)