Collaborative Recommendation with Multiclass Preference Context

Weike Pan and Zhong Ming*

{panweike,mingz}@szu.edu.cn
College of Computer Science and Software Engineering
Shenzhen University

Problem Definition

users \ items	1	2	3	4	5	6
1	?	3	?	?	?	?
2	?	?	3	1	4	?
3	?	2	?	1	4	?
4	?	?	?	?	2	5
5	?	5	?	5	?	5
6	4	?	?	1	?	?

We have n users (or rows) and m items (or columns), and some observed multiclass preferences such as ratings that are recorded in $\mathcal{R} = \{(u, i, r_{ui})\}$ with $r_{ui} \in \mathbb{M}$, where $\mathbb{M}$ can be $\{1, 2, 3, 4, 5\}$, $\{0.5, 1, 1.5, \ldots, 5\}$ or other ranges.

Our goal is to build a model so that the missing entries of the original matrix can be predicted.

Motivation

- Factorization- and neighborhood-based methods have been recognized as the state-of-the-art methods for collaborative recommendation tasks, e.g., rating prediction.
- Those two methods are known complementary to each other, while very few works have been proposed to combine them together.
- SVD++ tries to combine the main idea of latent features and neighborhood of those two methods, but ignores the existent *categorical scores* of the rated items.
- In this paper, we address this limitation of SVD++.

Overall of Our Solution

users \ items	1	2	3	4	5	6
1	?	3	?	?	?	?
2	?	?	3	<u>1</u>	<u>4</u>	?
3	?	2	?	1	4	?
4	?	?	?	?	2	5
5	?	5	?	5	?	5
6	4	?	?	1	?	?

$P(r_{23}|(2,3))$ MF

$P(r_{23}|(2,3);(2,4),(2,5))$ MF-OPC

$P(r_{23}|(2,3);(2,4,\underline{1}),(2,5,\underline{4}))$ MF-MPC

Preference Generalization Probability

Matrix Factorization with Multiclass Preference Context (MF-MPC)

- We take a user's ratings as categorical *multiclass preferences*.
- We integrate an enhanced neighborhood based on the assumption that users with similar past multiclass preferences (instead of oneclass preferences in MF-OPC, i.e., SVD++) will have similar taste in the future.

Advantage of Our Solution

- MF-MPC is able to make use of the multiclass preference context in the factorization framework in a fine-grained manner and thus inherits the advantages of factorization- and neighborhood-based methods in a better way.

Notations

Table: Some notations.

n	user number
m	item number
u, u'	user ID
i, i'	item ID
$\mathbb{M}$	multiclass preference set
$r_{ui} \in \mathbb{M}$	rating of user u on item i
$\mathcal{R} = \{(u, i, r_{ui})\}$	rating records of training data
$y_{ui} \in \{0, 1\}$	indicator, $y_{ui} = 1$ if $(u, i, r_{ui}) \in \mathcal{R}$
$\mathcal{I}_u^r, r \in \mathbb{M}$	items rated by user u with rating r
$\mathcal{I}_u$	items rated by user u
$\mu \in \mathbb{R}$	global average rating value
$b_u \in \mathbb{R}$	user bias
$b_i \in \mathbb{R}$	item bias
$d \in \mathbb{R}$	number of latent dimensions
$U_{u\cdot} \in \mathbb{R}^{1 \times d}$	user-specific latent feature vector
$V_{i\cdot}, O_{i\cdot}, M_{i\cdot}^r \in \mathbb{R}^{1 \times d}$	item-specific latent feature vector
$\mathcal{R}^{te} = \{(u, i, r_{ui})\}$	rating records of test data
$\hat{r}_{ui}$	predicted rating of user u on item i
T	iteration number in the algorithm

Preference Generalization Probability of MF

For a traditional matrix factorization (MF) model, the rating of user u on item i, r_{ui}, is assumed to be dependent on latent features of user u and item i only. We can represent it in a probabilistic way as follows,

$$P(r_{ui}|(u,i)), \tag{1}$$

which means that the probability of generating the rating r_{ui} is conditioned on the (user, item) pair (u,i) or their latent features only.

Prediction Rule of MF-OPC

Some advanced models assume that the rating r_{ui} is related to not only the user u and item i but also the other rated items by user u as a certain *context*, denoted as $\mathcal{I}_u \backslash \{i\}$. Similarly, the preference generalization probability can be represented as follows,

$$P(r_{ui}|(u,i); (u,i'), i' \in \mathcal{I}_u \backslash \{i\}), \tag{2}$$

where both (u,i) and $(u,i'), i' \in \mathcal{I}_u \backslash \{i\}$ denote the factors that govern the generalization of the rating r_{ui}. The advantage of the conditional probability in Eq.(2) is its ability to allow users with similar rated item sets to have similar latent features in the learned model. However, the exact values of the ratings assigned by the user u have not been exploited yet. Hence, we call the condition $(u,i'), i' \in \mathcal{I}_u \backslash \{i\}$ in Eq.(2) *oneclass* preference context (OPC).

Prediction Rule of MF-MPC

We go one step beyond and propose a fine-grained preference generalization probability,

$$P(r_{ui}|(u,i);(u,i',r_{ui'}),i' \in \cup_{r\in\mathbb{M}}\mathcal{I}_u^r\backslash\{i\}), \quad (3)$$

which includes the rating $r_{ui'}$ of each rated item by user u. This new probability is based on three parts, including (i) the (user, item) pair (u,i) in Eq.(1), (ii) the examined items $\cup_{r\in\mathbb{M}}\mathcal{I}_u^r\backslash\{i\}$ in Eq.(2), and (iii) the categorical score $r_{ui'}$ of each rated item.
The difference between the oneclass preference context $(u,i'),i' \in \mathcal{I}_u\backslash\{i\}$ in Eq.(2) and the condition $(u,i',r_{ui'}),i' \in \cup_{r\in\mathbb{M}}\mathcal{I}_u^r\backslash\{i\}$ in Eq.(3) is the categorical multiclass scores (or ratings), $r_{ui'}$, and thus we call it *multiclass* preference context (MPC).

Prediction Rule of MF

For a basic matrix factorization model, the prediction rule of the rating assigned by user u to item i is defined as follows,

$$\hat{r}_{ui} = U_{u\cdot} V_{i\cdot}^T + b_u + b_i + \mu, \quad (4)$$

where $U_{u\cdot} \in \mathbb{R}^{1\times d}$ and $V_{i\cdot} \in \mathbb{R}^{1\times d}$ are the user-specific and item-specific latent feature vectors, respectively, and b_u, b_i and μ are the user bias, the item bias and the global average, respectively.

Prediction Rule of MF-OPC

For matrix factorization with oneclass preference context, we can define the prediction rule of a rating as follows,

$$\hat{r}_{ui} = U_{u\cdot}V_{i\cdot}^T + \bar{U}_{u\cdot}^{\text{OPC}}V_{i\cdot}^T + b_u + b_i + \mu, \quad (5)$$

where $\bar{U}_{u\cdot}^{\text{OPC}}$ is based on the corresponding oneclass preference context $\mathcal{I}_u\backslash\{i\}$,

$$\bar{U}_{u\cdot}^{\text{OPC}} = \frac{1}{\sqrt{|\mathcal{I}_u\backslash\{i\}|}} \sum_{i'\in\mathcal{I}_u\backslash\{i\}} O_{i'\cdot}. \quad (6)$$

From the definition of $\bar{U}_{u\cdot}^{\text{OPC}}$ in Eq.(6), we can see that two users, u and u', with similar examined item sets, $\mathcal{I}_u$ and $\mathcal{I}_{u'}$, will have similar latent representations $\bar{U}_{u\cdot}^{\text{OPC}}$ and $\bar{U}_{u'\cdot}^{\text{OPC}}$. Hence, the prediction rule in Eq.(5) can be used to integrate certain neighborhood information.

Prediction Rule of MF-MPC

In matrix factorization with multiclass preference context, we propose a novel and generic prediction rule for the rating of user u to item i,

$$\hat{r}_{ui} = U_{u\cdot}V_{i\cdot}^T + \bar{U}_{u\cdot}^{\text{MPC}}V_{i\cdot}^T + b_u + b_i + \mu, \tag{7}$$

where $\bar{U}_{u\cdot}^{\text{MPC}}$ is from the multiclass preference context,,

$$\bar{U}_{u\cdot}^{\text{MPC}} = \sum_{r\in\mathbb{M}} \frac{1}{\sqrt{|\mathcal{I}_u^r\backslash\{i\}|}} \sum_{i'\in\mathcal{I}_u^r\backslash\{i\}} M_{i'\cdot}^r. \tag{8}$$

We can see that $\bar{U}_{u\cdot}^{\text{MPC}}$ in Eq.(8) is different from $\bar{U}_{u\cdot}^{\text{OPC}}$ in Eq.(6), because it contains more information, i.e., the fine-grained categorical preference of each rated item.

Objective Function of MF-MPC

With the prediction rule in Eq.(7), we can learn the model parameters in the following minimization problem,

$$\min_{\Theta} \sum_{u=1}^{n} \sum_{i=1}^{m} y_{ui} [\frac{1}{2}(r_{ui} - \hat{r}_{ui})^2 + reg(u,i)] \tag{9}$$

where $y_{ui} \in \{0,1\}$ is an indicator variable denoting whether (u, i, r_{ui}) is in the set of rating records $\mathcal{R}$, $reg(u,i) = \frac{\lambda}{2}||U_{u\cdot}||^2 + \frac{\lambda}{2}||V_{i\cdot}||^2 + \frac{\lambda}{2}||b_u||^2 + \frac{\lambda}{2}||b_i||^2 + \frac{\lambda}{2}\sum_{r\in\mathbb{M}}\sum_{i'\in\mathcal{I}_u^r\backslash\{i\}} ||M_{i'\cdot}^r||_F^2$ is the regularization term used to avoid overfitting, and $\Theta = \{U_{u\cdot}, V_{i\cdot}, b_u, b_i, \mu, M_{i\cdot}^r\}$, $u = 1, 2 \ldots, n$, $i = 1, 2, \ldots, m$, $r \in \mathbb{M}$ are model parameters to be learned. Note that the form of the objective function in Eq.(9) is exactly the same with that of the basic matrix factorization, because our improvement is reflected in the prediction rule for $\hat{r}_{ui}$.

Gradients of MF-MPC

For a tentative objective function $\frac{1}{2}(r_{ui} - \hat{r}_{ui})^2 + reg(u, i)$, we have the gradients of the model parameters,

$$\nabla U_{u\cdot} = -e_{ui} V_{i\cdot} + \lambda U_{u\cdot} \tag{10}$$

$$\nabla V_{i\cdot} = -e_{ui}(U_{u\cdot} + \bar{U}_{u\cdot}^{\text{MPC}}) + \lambda V_{i\cdot} \tag{11}$$

$$\nabla b_u = -e_{ui} + \lambda b_u \tag{12}$$

$$\nabla b_i = -e_{ui} + \lambda b_i \tag{13}$$

$$\nabla \mu = -e_{ui} \tag{14}$$

$$\nabla M_{i'\cdot}^r = \frac{-e_{ui} V_{i\cdot}}{\sqrt{|\mathcal{I}_u^r \backslash \{i\}|}} + \lambda M_{i'\cdot}^r, i' \in \mathcal{I}_u^r \backslash \{i\}, r \in \mathbb{M} \tag{15}$$

where $e_{ui} = (r_{ui} - \hat{r}_{ui})$ is the difference between the true rating and the predicted rating.

Update Rules of MF-MPC

Finally, we have the update rules,

$$\theta = \theta - \gamma \nabla \theta, \tag{16}$$

where γ is the learning rate, and $\theta \in \Theta$ is a model parameter to be learned.

Algorithm of MF-MPC

```
Initialize model parameters Θ
for t = 1, ..., T do
    for t_2 = 1, ..., |R| do
        Randomly pick up a rating from R
        Calculate the gradients via Eq.(10-15)
        Update the parameters via Eq.(16)
    end for
    Decrease the learning rate γ ← γ × 0.9
end for
```

Figure: The algorithm of MF-MPC.

Analysis

- MPC in Eq.(3) will be reduced to the OPC in Eq.(2) when we treat all ratings as a constant.
- Hence, SVD++ is a special case of our MF-MPC.

Detasets

- MovieLens100K (i.e., ML100K): 100000 ratings by 943 users and 1682 items, $\mathbb{M} = \{1, 2, 3, 4, 5\}$
- MovieLens1M (i.e., ML1M): 1000209 ratings by 6040 users and 3952 items, $\mathbb{M} = \{1, 2, 3, 4, 5\}$
- MovieLens10M (i.e., ML10M): 10000054 ratings by 71567 users and 10681 items, $\mathbb{M} = \{0.5, 1, 1.5, \ldots, 5\}$

For each data set, we first divide it into five parts with equal size. Then, we take one part as test data and the remaining four parts as training data, which is repeated for five times so that we have five copies of training data and test data for each of the three data sets.
The averaged rating prediction performance on those five copies of test data will be reported.

Baselines

- AF (average filling): we use the average rating of each user as calculated from the training data $\mathcal{R}$ to predict each rating in the test data;
- CF (collaborative filtering): we implement a user-oriented neighborhood-based collaborative filtering method using PCC (Pearson correlation coefficient) as the similarity measurement;
- MF (matrix factorization): we use the basic latent factor model, i.e., matrix factorization without preference context as shown in Eq.(4), as a major baseline; and
- MF-OPC (matrix factorization with oneclass preference context): for direct comparative studies between MPC and OPC, we also use MF-OPC as shown in Eq.(5). Note that MF-OPC is the same with SVD++.

Initialization of Model Parameters

We use the statistics of training data $\mathcal{R}$ to initialize the model parameters,

$$
\begin{aligned}
\mu &= \sum_{u=1}^{n}\sum_{i=1}^{m} y_{ui} r_{ui} / \sum_{u=1}^{n}\sum_{i=1}^{m} y_{ui} \\
b_u &= \sum_{i=1}^{m} y_{ui}(r_{ui} - \mu) / \sum_{i=1}^{m} y_{ui} \\
b_i &= \sum_{u=1}^{n} y_{ui}(r_{ui} - \mu) / \sum_{u=1}^{n} y_{ui} \\
U_{uk} &= (r - 0.5) \times 0.01, k = 1, \ldots, d \\
V_{ik} &= (r - 0.5) \times 0.01, k = 1, \ldots, d \\
M^{r}_{i'k} &= (r - 0.5) \times 0.01, k = 1, \ldots, d
\end{aligned}
$$

where r $(0 \leq r < 1)$ is a random variable.

Parameter Configurations

For factorization-based methods:

- The learning rate γ is initialized as 0.01
- The number of latent dimensions: $d = 20$
- Iteration number: $T = 50$
- We search the best value of the tradeoff parameter λ from $\{0.001, 0.01, 0.1\}$ using the first copy of each data and the RMSE metric

For neighborhood-based method (i.e., CF), we set it to be the same with the value of d, i.e., 20.
We also use different dimensions for MF and MF-OPC in order to study the effectiveness of our MF-MPC from the perspective of the number of model parameters.

Initialization of Model Parameters

We use the statistics of training data $\mathcal{R}$ to initialize the model parameters,

$$
\begin{aligned}
\mu &= \sum_{u=1}^{n}\sum_{i=1}^{m} y_{ui} r_{ui} / \sum_{u=1}^{n}\sum_{i=1}^{m} y_{ui} \\
b_u &= \sum_{i=1}^{m} y_{ui}(r_{ui} - \mu) / \sum_{i=1}^{m} y_{ui} \\
b_i &= \sum_{u=1}^{n} y_{ui}(r_{ui} - \mu) / \sum_{u=1}^{n} y_{ui} \\
U_{uk} &= (r - 0.5) \times 0.01, k = 1, \ldots, d \\
V_{ik} &= (r - 0.5) \times 0.01, k = 1, \ldots, d \\
M_{i'k}^{p} &= (r - 0.5) \times 0.01, k = 1, \ldots, d
\end{aligned}
$$

where r $(0 \leq r < 1)$ is a random variable.

Post-Processing

- For the rating range (grade score set) $\mathbb{M} = \{1, 2, 3, 4, 5\}$
  If $\hat{r}_{ui} > 5$, $\hat{r}_{ui} = 5$
  If $\hat{r}_{ui} < 1$, $\hat{r}_{ui} = 1$
- For the rating range (grade score set) $\mathbb{M} = \{0.5, 1, 1.5, \dots, , 5\}$
  If $\hat{r}_{ui} > 5$, $\hat{r}_{ui} = 5$
  If $\hat{r}_{ui} < 0.5$, $\hat{r}_{ui} = 0.5$

Evaluation Metrics

- Mean Absolute Error (MAE)

$$MAE = \sum_{(u,i,r_{ui}) \in \mathcal{R}^{te}} |r_{ui} - \hat{r}_{ui}| / |\mathcal{R}^{te}|$$

- Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\sum_{(u,i,r_{ui}) \in \mathcal{R}^{te}} (r_{ui} - \hat{r}_{ui})^2 / |\mathcal{R}^{te}|}$$

- Performance: the smaller the better.

Results (1/2)

Table: Recommendation performance on MAE and RMSE ($d = 20$).

Data	Metric	AF	CF	MF	MF-OPC/SVD++	MF-MPC
ML100K	MAE RMSE	0.8348±0.0025 1.0417±0.0019	0.7576±0.0028 0.9637±0.0039	0.7478±0.0032 0.9448±0.0030 ($\lambda = 0.1$)	0.7266±0.0032 0.9253±0.0032 ($\lambda = 0.001$)	**0.7092**±0.0032 **0.9091**±0.0026 ($\lambda = 0.001$)
ML1M	MAE RMSE	0.8289±0.0020 1.0355±0.0024	0.7560±0.0020 0.9531±0.0024	0.6956±0.0021 0.8832±0.0023 ($\lambda = 0.001$)	0.6655±0.0014 0.8511±0.0017 ($\lambda = 0.001$)	**0.6596**±0.0017 **0.8439**±0.0018 ($\lambda = 0.01$)
ML10M	MAE RMSE	0.7688±0.0007 0.9784±0.0008	0.7138±0.0003 0.9148±0.0005	0.6068±0.0006 0.7911±0.0008 ($\lambda = 0.01$)	0.6028±0.0003 0.7870±0.0006 ($\lambda = 0.01$)	**0.5947**±0.0003 **0.7783**±0.0005 ($\lambda = 0.01$)

Observations:

- MF-MPC performs significantly better than all baselines on all three data sets, which clearly shows the effectiveness of our proposed multiclass preference context;
- MF-OPC or SVD++ is much better than MF, which shows the complementarity of factorization-based methods and neighborhood-based methods.

Results (2/2)

Table: Recommendation performance on RMSE ($d = 20$ for MF-MPC, $d = 120$ for MF and $d = 80$ for MF-OPC).

Data	MF	MF-OPC	MF-MPC
ML100K	0.9439$\pm$0.0036 ($\lambda = 0.001$)	0.9209$\pm$0.0034 ($\lambda = 0.001$)	**0.9091**$\pm$0.0026 ($\lambda = 0.001$)
ML1M	0.8719$\pm$0.0023 ($\lambda = 0.001$)	0.8477$\pm$0.0017 ($\lambda = 0.001$)	**0.8439**$\pm$0.0018 ($\lambda = 0.01$)
ML10M	0.7821$\pm$0.0005 ($\lambda = 0.01$)	0.7810$\pm$0.0005 ($\lambda = 0.01$)	**0.7783**$\pm$0.0005 ($\lambda = 0.01$)

Observations:

- MF-MPC does not have any advantage of using more model parameters for all three data sets.
- MF-MPC is again significantly better ($p < 0.01$) than MF and MF-OPC, which clearly shows the advantage of the proposed factorization model.

Conclusion

- We propose a novel method (i.e., MF-MPC) that integrates multiclass preference context (MPC) into the matrix factorization (MF) framework for rating prediction, and obtain significantly better results.
- We are interested in generalizing the idea of multiclass preference context to recommendation with categorical preference information in cross-domain scenarios.
- We are also interested in designing some advanced sampling strategy instead of the random sampling approach in the learning algorithm.

Thank you!

- We thank the handling editor and reviewers for their expert comments and constructive suggestions.
- We thank Yuchao Duan for some preliminary empirical studies and helpful discussions.
- We thank the support of National Natural Science Foundation of China No. 61502307 and No. 61672358, and Natural Science Foundation of Guangdong Province No. 2014A030310268 and No. 2016A030313038.