Federated Probabilistic Preference Distribution Modelling with Compactness Co-Clustering for Privacy-Preserving Multi-Domain Recommendation

Weiming Liu[1] , **Chaochao Chen**[1*] , **Xinting Liao**[1] , **Mengling Hu**[1] , **Jianwei Yin**[1] , **Yanchao Tan**[2] and **Longfei Zheng**[3]
[1]College of Computer Science and Technology, Zhejiang University, Hangzhou, China
[2]College of Computer and Data Science, Fuzhou University, Fuzhou, China
[3]Ant Financial, Hangzhou, China
{21831010, zjuccc, xintingliao, humengling, zjuyjw}@zju.edu.cn, yctan@fzu.edu.cn, zlf206411@antfin.com

Abstract

With the development of modern internet techniques, Cross-Domain Recommendation (CDR) systems have been widely exploited for tackling the data-sparsity problem. Meanwhile most current CDR models assume that user-item interactions are accessible across different domains. However, such knowledge sharing process will break the privacy protection policy. In this paper, we focus on the Privacy-Preserving Multi-Domain Recommendation problem (PPMDR). The problem is challenging since different domains are sparse and heterogeneous with the privacy protection. To tackle the above issues, we propose Federated Probabilistic Preference Distribution Modelling (`FPPDM`). `FPPDM` includes two main components, i.e., *local domain modelling component* and *global server aggregation component* with federated learning strategy. The local domain modelling component aims to exploit user/item preference distributions using the rating information in the corresponding domain. The global server aggregation component is set to combine user characteristics across domains. To better extract semantic neighbors information among the users, we further provide compactness co-clustering strategy in `FPPDM++` to cluster the users with similar characteristics. Our empirical studies on benchmark datasets demonstrate that `FPPDM`/`FPPDM++` significantly outperforms the state-of-the-art models.

1 Introduction

Cross-domain recommendation (CDR) systems have shown great success in alleviating the serious data sparsity problem [Guo *et al.*, 2021; Li *et al.*, 2022; Zang *et al.*, 2021a; Zheng *et al.*, 2022]. It is well acknowledged that CDR models always share the user-item rating information across domains for joint collaborative filtering. Such methods can provide better performance on both source and target domains. However, with the increasing law on privacy protection, the user-item rating information are not accessible among different domains [Chen and et al, 2022]. How to provide high-quality cross- and multi-domain recommendations under the demand of privacy protection has become an urgent problem.

In this paper, we focus on a more general problem of Privacy-Preserving Multi-Domain Recommendation (PPMDR). That is, user-item rating interactions are regarded as private information and other domains (participants) cannot directly obtain them. Such demand makes it difficult for knowledge transfer across domains to extract users' domain-invariant preferences. Meanwhile, these multiple domains all suffer from the data sparsity problem. It further hurdles the model to learning reliable representations and limits the model's performance.

Existing research on CDR cannot solve the PPMDR problem well. Firstly, most conventional CDR models need to access the whole user-item rating interactions among different domains for modelling [Hu *et al.*, 2018; Chen *et al.*, 2020]. While it is not suitable under the settings of PPMDR, since the user-item ratings should be protected in each domain. Recently, FedMF [Chai *et al.*, 2020] first adopts federated learning strategy for ensuring data privacy protection. However, FedMF only adopted a shallow model and thus cannot depict the complex user-item relationship well. Secondly, most of current CDR models only utilize embeddings rather than distributions to represent users and items, which leads to an inaccurate understanding of user preference, especially under the data sparsity scenario [Jiang *et al.*, 2020; Ma and et al, 2020]. Specifically, we provide an example in Fig.1(a)-(b) where the user Mary prefers romantic and historical items while dislikes horror items. If we only consider the user/item embeddings, the horror items *Suspect X* or *Kill Game* will be recommended to Mary since they are closer than *Woman in Love* or *Flipped*. While when we model the users/items with Gaussian distributions, the users/items can represent more richer meanings. As depicted in Fig.1(a), Mary's distribution overlaps her prefer books (e.g., *Woman in Love* and *Global History*) rather than horror items to avoid inaccurate results. While only using single domain information cannot better model users' preference distribution, as shown in Fig.1(b), Mary's local distribution with dashed eclipse does not overlap with the romantic movie *Titanic* that

*Chaochao Chen is the corresponding author.

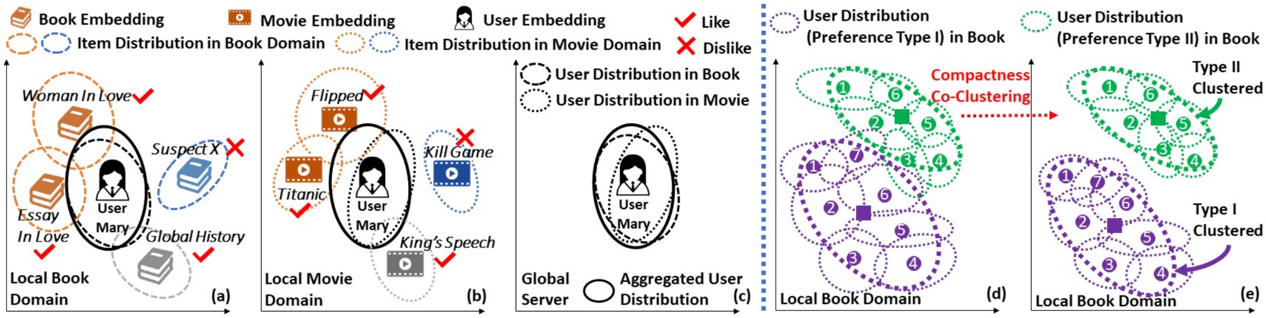

Figure 1: The illustrations of probabilistic preference distribution modelling with compactness co-clustering.

she likes. Then we should adopt the global server to aggregate the user distributions for knowledge sharing in Fig.1(c). Meanwhile, different domains are always heterogeneous with diverse kinds of items. Hence, it could be rather difficult to directly apply the commonly-used federated learning methods for information gathering. What is more, most of the CDR models only concentrate on user-item interactions and neglect potential semantic relations. As shown in Fig.1(d)-(e), the user preference distributions are scattered and users with different tastes still have overlapped space. In conclusion, how to properly model and aggregate the user/item preference distribution is still a challenging problem.

To address the aforementioned issues, in this paper, we propose Federated Probabilistic Preference Distribution Modelling (`FPPDM`) for solving the PPMDR problem. `FPPDM` involves two main components, i.e., *local domain modelling component* and *global server aggregation component* where user-item rating interactions are stored in the local domain due to the demand for privacy protection. We further propose neural networks to capture user/item preference distribution via modelling the corresponding mean and covariance in the local domain modelling component. Then we aggregate these users' preference distributions in the global server aggregation component and send them back to local domains. To further leverage the user similarity relationship, we propose a compactness co-clustering method in `FPPDM++`. The compactness co-clustering method can gather users based on their tastes or characteristics to provide more satisfactory results. We summarize our main contributions as follows: (1) We propose a novel federal learning framework, i.e., `FPPDM`, for solving the PPMDR problem and protecting the privacy of user-item rating information. `FPPDM` involves multiple local domains and a global server to model and aggregate user/item preference distributions. (2) To achieve better results, we further propose a compactness co-clustering method to leverage similar users' information in `FPPDM++`. (3) Extensive empirical studies on Douban and Amazon datasets demonstrate that both `FPPDM` and `FPPDM++` significantly improve the state-of-the-art models, especially under the PPMDR setting.

2 Related Work

Cross- and Multi-Domain Recommendation. The existing CDR models have two main types, i.e., *single-target CDR* models and *dual-target CDR* models [Zang *et al.*, 2021b; Zhu *et al.*, 2021; Zhang *et al.*, 2022b; Yuan *et al.*, 2019; Wang and et al, 2022; Sun and et al, 2022]. Single-target CDR models leverage the information from the relevant source domain with rich data to the sparse target domain. Conventional single-target CDR models aims to adopt matrix factorization mechanism for collaborative filtering on overlapped users/items [Lian *et al.*, 2017]. More recently, some single-target CDR models also adopt the deep neural networks to users/items preferences [Elkahky *et al.*, 2015; Hu *et al.*, 2018]. Latest PriCDR [Chen and et al, 2022] further considers the privacy protection when distilling the information from rich to sparse domains. Dual-target CDR models aims to simultaneously share and transfer information when both source and target domains are sparse. DOML [Li and Tuzhilin, 2021] first tried attempt to share dual knowledge transfer based on orthogonal mapping. Recently, researchers also utilized graph-based approaches [Zhao *et al.*, 2019; Liu *et al.*, 2020; Cao *et al.*, 2022] for solving dual-target CDR problem. Zhao et.al. [Zhao *et al.*, 2022] further enhanced the method with gate mechanism into multi-domain recommendation. While the above methods should distribute the local user-item ratings across different domains which could raise the concern on privacy protection. Hence they cannot be suitable for solving the PPMDR problem.

Federated Learning. With the increasing emphasis on protecting personal privacy, federated learning is widely used in diverse applications [Zhang *et al.*, 2022a; Meihan *et al.*, 2022]. FedAvg [McMahan *et al.*, 2017] was first be proposed for protecting the clients' privacy in distributed learning. To better satisfy the non-identical independent fashion, proximal regularization term [T Dinh *et al.*, 2020] or primal-dual algorithm [Zhang *et al.*, 2021; Acar *et al.*, 2021] have been added in the local clients. Furthermore, shared representations have been exploited [Pillutla *et al.*, 2022] to provide personal solutions. Meanwhile knowledge distill method [Lin *et al.*, 2020] and prototype alignment [Tan and et al, 2021; Dong *et al.*, 2022] have been proposed with the consideration of heterogeneous data distribution or network architecture. More recently, FedMF [Chai *et al.*, 2020] successfully adopted the federated learning into the field of cross domain recommendation. Latest FedCDR [Meihan *et al.*, 2022] also adopted federated learning into solving cold-start CDR problem. However, FedMF only adopted shallow model which cannot capture and depict the complex and complicate user-item relations. While FedCDR separated the rating prediction stage and transfer stage which limits the model performance. As a comparison, in this paper, we extend the federated learning with PPMDR problem by adopting distribution aggregation method with compactness co-clustering strategy.

3 Methodology

First, we describe notations. We assume there are K ($K \geq 2$) domains (clients) as $\{\mathbb{D}^{(1)}, \mathbb{D}^{(2)}, \cdots, \mathbb{D}^{(K)}\}$ where $\mathbb{D}^{(i)}$ denotes as the i-th domain (client). We assume each domain share the same set of N_U users. Meanwhile each domain has $N_V^{(i)}$ different types of items. Let $\boldsymbol{R}^{(i)} \in \mathbb{R}^{N_U \times N_V^{(i)}}$ be the observed rating matrices in i-th domain $\mathbb{D}^{(i)}$ and users in different domains are overlapped. To simplify the problem, in this paper, we assume both domains have no other auxiliary information. Meanwhile, the rating matrix $\boldsymbol{R}^{(i)}$ is the private data which cannot be directly shared in the public space.

Then, we introduce the overview of our proposed FPPDM framework, as is illustrated in Fig. 2. FPPDM mainly involves two components, i.e., *local domain modelling component* and *global server aggregation component*. The i-th local domain stores the user-item rating $\boldsymbol{R}^{(i)}$ in i-th domain $\mathbb{D}^{(i)}$. The local domain aims to model user/item probabilistic preference distribution based on the rating interactions. The global server aggregation component aims to gather the overlapped user information for knowledge sharing. To achieve more better recommendation results, we further provide FPPDM++ with compactness co-clustering method. This method tends to exploit and cluster users with similar tastes and characteristics.

3.1 The Framework of FPPDM

Local Domain Modelling Component

Firstly, we provide the details of the local domain module in FPPDM. For the i-th user and the j-th item in the $\mathbb{D}^{(k)}$ domain, we define their corresponding one-hot ID vectors as $\boldsymbol{X}_i^{U_{(k)}}$ and $\boldsymbol{X}_j^{V_{(k)}}$, respectively. We adopt a trainable lookup table to exploit the user/item one-hot ID embedding as $\boldsymbol{E}_i^{U_{(k)}} = \mathrm{LookUp}(\boldsymbol{X}_i^{U_{(k)}})$ and $\boldsymbol{E}_j^{V_{(k)}} = \mathrm{LookUp}(\boldsymbol{X}_j^{V_{(k)}})$ respectively. To better aggregate useful information among the user-item interactions, we further adopt the commonly-used graph neural network in modelling. Then we first build up the corresponding interaction $\boldsymbol{A}^{(k)}$ graph among users and items as $\boldsymbol{A}^{(k)} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{R}^{(k)} \\ (\boldsymbol{R}^{(k)})^\top & \boldsymbol{0} \end{bmatrix}$. Then we can conduct the graph convolution neural network for modelling user/item preference distribution:

$$\begin{aligned} [\boldsymbol{\mu}^{U_{(k)}}, \boldsymbol{\mu}^{V_{(k)}}] &= \mathrm{GCN}(\cdots \mathrm{GCN}(\boldsymbol{E}^{(k)}, \boldsymbol{A}^{(k)} | \boldsymbol{W}_\mu^{(k)}) \cdots), \\ [(\boldsymbol{\sigma}^{U_{(k)}})^2, (\boldsymbol{\sigma}^{V_{(k)}})^2] &= \mathrm{GCN}(\cdots \mathrm{GCN}(\boldsymbol{E}^{(k)}, \boldsymbol{A}^{(k)} | \boldsymbol{W}_\sigma^{(k)}) \cdots), \end{aligned} \tag{1}$$

where $\boldsymbol{E}^{(k)} = \{\boldsymbol{E}^{U_{(k)}}, \boldsymbol{E}^{V_{(k)}}\}$ denotes the set of users and items in $\mathbb{D}^{(k)}$. The $\boldsymbol{W}_\mu^{(k)}$ and $\boldsymbol{W}_\sigma^{(k)}$ denote the trainable weights. $\mathrm{GCN}(\cdot)$ denotes the graph convolution network operation which can be computed as:

$$\mathrm{GCN}(\boldsymbol{E}^{(k)}, \boldsymbol{A}^{(k)} | \boldsymbol{W}^{(k)}) = (\widetilde{\boldsymbol{D}}^{(k)})^{-\frac{1}{2}} \widetilde{\boldsymbol{A}}^{(k)} (\widetilde{\boldsymbol{D}}^{(k)})^{-\frac{1}{2}} \boldsymbol{E}^{(k)} \boldsymbol{W}^{(k)}, \tag{2}$$

where $\widetilde{\boldsymbol{D}}^{(k)} = \mathrm{diag}(\widetilde{\boldsymbol{A}}^{(k)} \boldsymbol{1})$ denotes the degree matrix for the graph $\widetilde{\boldsymbol{A}}^{(k)}$ and $\widetilde{\boldsymbol{A}}^{(k)} = \boldsymbol{A}^{(k)} + \boldsymbol{I}$. Specifically, we adopt ℓ-th layers of graph convolution network layers to achieve the users/items' mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\sigma}^2$ of their embeddings distribution as follows:

$$\mathbb{P}(\boldsymbol{U}^{(k)}) = \mathcal{N}(\boldsymbol{\mu}^{U_{(k)}}, (\boldsymbol{\sigma}^{U_{(k)}})^2), \ \mathbb{P}(\boldsymbol{V}^{(k)}) = \mathcal{N}(\boldsymbol{\mu}^{V_{(k)}}, (\boldsymbol{\sigma}^{V_{(k)}})^2), \tag{3}$$

where $\mathbb{P}(\boldsymbol{U}^{(k)})$ and $\mathbb{P}(\boldsymbol{V}^{(k)})$ denote the local user and item distributions respectively. Since using the single user or item embeddings cannot precisely depict user-item relationship, we adopt the Gaussian distribution to parameterize user and item distribution. Specifically, the Gaussian distribution can capture the complex and complicated user/item preference via measuring the covariance. After that we should train the model to fit the observed user-item ratings. To better achieve this goal, we propose the distribution-based metric learning loss as given below:

$$\ell_R^{(k)} = - \sum_{(\boldsymbol{U}_i^{(k)}, \boldsymbol{V}_j^{(k)}) \in \mathcal{O}^{(k)}} \log \frac{e^{-\mathcal{D}(\mathbb{P}(\boldsymbol{U}_i^{(k)}), \mathbb{P}(\boldsymbol{V}_j^{(k)}))}}{\sum_{\boldsymbol{V}_{\mathrm{neg},(i)}^{(k)}} e^{-\mathcal{D}(\mathbb{P}(\boldsymbol{U}_i^{(k)}), \mathbb{P}(\boldsymbol{V}_{\mathrm{neg},(i)}^{(k)}))}}, \tag{4}$$

where $\mathcal{O}^{(k)}$ denotes the positive user-item pairs. $\boldsymbol{V}_{\mathrm{neg},(i)}^{(k)}$ denotes the negative items for the i-th user. The $\mathcal{D}(\cdot)$ denotes Wasserstein distance among different Gaussian distributions which can be calculated as:

$$\mathcal{D}(\mathbb{P}(\boldsymbol{U}_i^{(k)}), \mathbb{P}(\boldsymbol{V}_j^{(k)})) = ||\boldsymbol{\mu}_i^{U_{(k)}} - \boldsymbol{\mu}_j^{V_{(k)}}||_2^2 + ||\boldsymbol{\sigma}_i^{U_{(k)}} - \boldsymbol{\sigma}_j^{V_{(k)}}||_2^2.$$

After adopting the metric-based rating prediction loss, we can pull the positive user-item pairs while push away the negative user-item pairs. Meanwhile, the local and global overlapped user distributions should be consistent for knowledge sharing. To this end, we propose a regularization term to reduce the distance among these corresponding local and global user distributions as follows:

$$\ell_P^{(k)} = \Sigma_{i=1}^N [||\boldsymbol{\mu}_i^{U_{(k)}} - \widehat{\boldsymbol{\mu}}_i||_2^2 + ||\boldsymbol{\sigma}_i^{U_{(k)}} - \widehat{\boldsymbol{\sigma}}_i||_2^2], \tag{5}$$

where N denotes as the batchsize. $\mathbb{P}(\widehat{\boldsymbol{U}}_i) = \mathcal{N}(\widehat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\sigma}}_i^2)$ denotes the i-th global user distribution. By combining the rating prediction loss $L_R^{(k)}$ with the regularization term $L_P^{(k)}$, we can obtain the total loss in local domain modelling component:

$$\min \ell_{\mathrm{FPPDM}} = \ell_R^{(k)} + \lambda_P \ell_P^{(k)}, \tag{6}$$

where λ_P denotes as the balanced hyper parameter. After the training in local domains, we can obtain the local user distribution $\mathbb{P}(\boldsymbol{U}^{(k)})$ and send them to the global server.

Global Server Aggregation Component

Since the global server have received the local user distributions from multiple different domains, the global server should aggregate and update the global user distributions $\mathbb{P}(\widehat{\boldsymbol{U}})$. This process is important since the data sparsity problem will deteriorate the user-item modelling in local domain. As an example in Fig.1(b), the local user distribution of Mary will not fully cover her interests since it still away from the romantic movie *Titanic*. Then we adopt the server aggregation

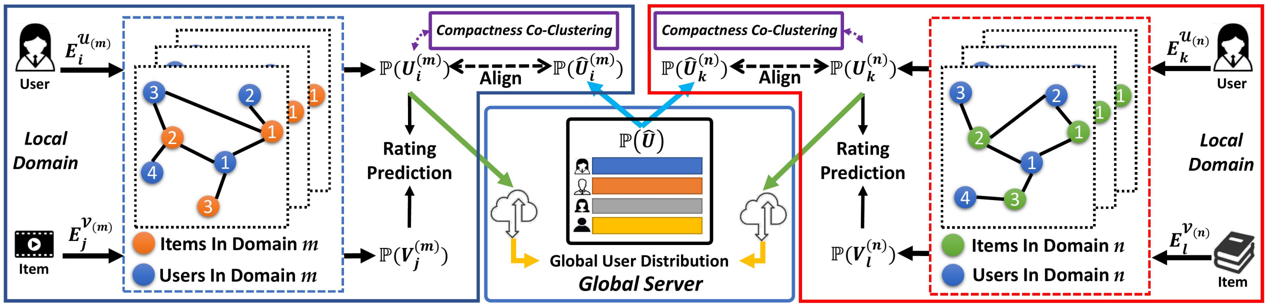

Figure 2: The model framework of FPPDM with includes the multiple local domains and global server.

to obtain the global user preference distribution in Fig.1(c) and after that it will enhance the model performance in the local movie domain in Fig.1(b). To fulfill this task, we tend to exploit the new global user distributions $\mathbb{P}(\widehat{U})$ which has the smaller Wasserstein distance among these local user distributions as $[\min_{\mathbb{P}(\widehat{U})} \frac{1}{K}\sum_{i=1}^{N}\sum_{k=1}^{K}\mathcal{D}(\mathbb{P}(\widehat{U}_i), \mathbb{P}(U_i^{(k)}))]$. By taking the differentiation w.r.t. on $\widehat{\mu}_i$ and $\widehat{\sigma}_i$, we can easily obtain the optimal results as:

$$\widehat{\mu}_i = \frac{\Sigma_{k=1}^{K}\mu_i^{U_{(k)}}}{K}, \quad \widehat{\sigma}_i = \frac{\Sigma_{k=1}^{K}\sigma_i^{U_{(k)}}}{K}. \tag{7}$$

After we achieve the aggregated global user distributions, we further send them back to the local domains. Meanwhile only mean and covariance of user preference distributions will be transmitted to the global server which makes it different from conventional federated learning methods (e.g., FedAvg [McMahan *et al.*, 2017]). Note that it is difficult to reconstruct the raw user-item rating interactions from user preference distributions. Furthermore, one can even adopt encryption transportation method (e.g., homomorphic encryption [Gentry, 2009; Aono *et al.*, 2017]) on mean and covariance to strengthen security level.

3.2 The Framework of FPPDM++

Although FPPDM can share and transfer useful knowledge via the overlapped users across domains, the user preference distributions will become scattered in the latent space as shown in Fig.1(d). Specifically, the purple and green dots represent the users with different types of preferences in Fig.1(d). Some users (e.g., purple #7 and green #2) with different preferences still have overlapped distributions. The scattered latent distributions will hurdle the model performance by providing inaccurate results. The main reason behind is FPPDM cannot fully exploit and incorporate the semantic neighbors to enhance the model performance. That is, the users with similar tastes or preferences are their semantic neighbors who should be clustered. Meanwhile utilizing the co-clustering strategy can even filter out inherit data noise to further reduce data sparsity problem [Lin *et al.*, 2022b; Lin *et al.*, 2022a]. However, previous methods mainly focus on the simple embedding situation and they cannot be directly applied for distribution clustering. To resolve this issue, we first propose compactness co-clustering method in FPPDM++ to enhance the model performance by extracting useful semantic information among these preference distributions.

Algorithm 1 FPPDM / FPPDM++

```
Server executes:
Initialize global user distribution as P(Û).
for round = 1 to T do
    for each domain = k to K in parallel do
        Obtain P(U_i^(k)) ← LocalUpdate(i, P(Û))
    end for
    Update P(Û) via Eq.(7) and send to local domain.
end for
LocalUpdate(i, P(Û)):
for epoch = 1 to t do
    Calculate the rating prediction loss in Eq.(4).
    Calculate the regularization loss in Eq.(5).
    if The model is FPPDM++ then
        Optimize the compactness similarity in Eq.(8).
        Optimize the assign indication in Eq.(10).
        Calculate compactness co-clustering loss in Eq.(11).
    end if
    Update the local model.
end for
Return: Well-trained local model.
```

Then we introduce the details of our proposed compactness co-clustering method. To start with, we tend to figure out the similarity $s_{ij}^{(k)}$ between the i-th and j-th users in the $\mathbb{D}^{(k)}$ domain. Meanwhile we suppose to cluster the users into M groups to achieve more compact user representations. Specifically, we present the compactness similarity optimization problem as follows:

$$\begin{aligned} &\min_{s^{(k)}} \sum_{i,j=1}^{N} [s_{ij}^{(k)}\mathcal{D}(\mathbb{P}(U_i^{(k)}), \mathbb{P}(U_j^{(k)})) + \epsilon s_{ij}^{(k)} \log s_{ij}^{(k)}] \\ &s.t. \sum_{j=1}^{N} s_{ij}^{(k)} = 1, s_{ij}^{(k)} \geq 0, \mathrm{rank}(L^{(k)}) = N - M, \end{aligned} \tag{8}$$

where $L^{(k)} = \mathrm{diag}(S^{(k)}\mathbf{1}) - S^{(k)}$ denotes the Laplacian similarity matrix. ϵ denotes the hyper parameter to balance the similarity matching and the entropy regularization term $s_{ij}^{(k)} \log s_{ij}^{(k)}$. Meanwhile the entropy regularization term is set to obtain the nonnegative and nontrivial solution [Bai and Liang, 2020; Nie *et al.*, 2014]. Furthermore, we add the rank constraint, i.e., $\mathrm{rank}(L^{(k)}) = N - M$, on Laplacian similarity matrix $L^{(k)}$ to obtain more compact results by avoiding the situation that most users clustered in one group. However, it is difficult to directly optimize the rank constraint. Here, we make the approximation by calculating the minimal eigenvalues of $\boldsymbol{L}^{(k)}$ to replace the rank constraint. In other words, $\mathrm{rank}(\boldsymbol{L}^{(k)}) = N - M$ is equivalent to $\sum_{m=1}^{M} \phi_m(\boldsymbol{L}^{(k)})$ where $\phi_m(\boldsymbol{L}^{(k)})$ denotes the m-th smallest eigenvalue of $\boldsymbol{L}^{(k)}$. According to the Ky Fan's theorem [Fan, 1949], $\phi_m(\boldsymbol{L}^{(k)})$ can be calculated as:

$$\sum_{m=1}^{M} \phi_m(\boldsymbol{L}^{(k)}) = \min_{(\boldsymbol{F}^{(k)})^\top \boldsymbol{F}^{(k)} = \boldsymbol{I}} \mathrm{Tr}((\boldsymbol{F}^{(k)})^\top \boldsymbol{L}^{(k)} \boldsymbol{F}^{(k)}), \quad (9)$$

where $\boldsymbol{F}^{(k)} \in \mathbb{R}^{N \times M}$ is the indicator matrix. We provide the Lagrange multipliers for the original problem as follows:

$$\min_{\boldsymbol{S}^{(k)}, \boldsymbol{F}^{(k)}} J^{(k)} = \sum_{i,j=1}^{N} [s_{ij}^{(k)} \mathcal{D}(\mathbb{P}(\boldsymbol{U}_i^{(k)}), \mathbb{P}(\boldsymbol{U}_j^{(k)})) + \epsilon s_{ij}^{(k)} \log s_{ij}^{(k)}]$$
$$+ \eta \sum_{i,j=1}^{N} s_{ij}^{(k)} ||\boldsymbol{f}_i^{(k)} - \boldsymbol{f}_j^{(k)}||_2^2 + \sum_{i=1}^{N} \gamma_i^{(k)} \left(\sum_{j=1}^{N} s_{ij}^{(k)} - 1 \right),$$

where η is the hyper parameter and $\gamma_i^{(k)}$ is the multiplier. We further make simplification for rank constraint since $\sum_{i=1}^{N} \sum_{j=1}^{N} s_{ij}^{(k)} ||\boldsymbol{f}_i^{(k)} - \boldsymbol{f}_j^{(k)}||_2^2 = 2\mathrm{Tr}((\boldsymbol{F}^{(k)})^\top \boldsymbol{L}^{(k)} \boldsymbol{F}^{(k)})$. We first fix and $\boldsymbol{F}^{(k)}$ and update the $\boldsymbol{S}^{(k)}$ as:

$$s_{ij}^{(k)} = \frac{\exp(-(\mathcal{D}(\mathbb{P}(\boldsymbol{U}_i^{(k)}), \mathbb{P}(\boldsymbol{U}_j^{(k)})) + \eta ||\boldsymbol{f}_i^{(k)} - \boldsymbol{f}_j^{(k)}||_2^2)/\epsilon)}{\sum_{l=1}^{N} \exp(-(\mathcal{D}(\mathbb{P}(\boldsymbol{U}_i^{(k)}), \mathbb{P}(\boldsymbol{U}_l^{(k)})) + \eta ||\boldsymbol{f}_i^{(k)} - \boldsymbol{f}_l^{(k)}||_2^2)/\epsilon)}.$$

Then we fix $\boldsymbol{S}^{(k)}$ and update $\boldsymbol{F}^{(k)}$. The optimal solution $\boldsymbol{F}^{(k)}$ can be obtained by the M eigenvectors of $\boldsymbol{L}^{(k)}$ corresponding to the M smallest eigenvalues. We can alternatively update $\boldsymbol{S}^{(k)}$ and $\boldsymbol{F}^{(k)}$ to achieve the optimal solution.

Once we obtain the indicator matrix $\boldsymbol{F}^{(k)}$, we tend to figure out the relationship between each data and the corresponding clusters. To achieve this goal, we propose entropy-based K-Means method as follows:

$$\min_{\boldsymbol{\zeta}^{(k)} \mathbf{1} = 1, \boldsymbol{Z}^{(k)}} \sum_{i=1}^{N} \sum_{j=1}^{M} [\zeta_{ij}^{(k)} ||\boldsymbol{f}_i^{(k)} - \boldsymbol{z}_j^{(k)}||_2^2 + \epsilon \zeta_{ij}^{(k)} \log \zeta_{ij}^{(k)}], \quad (10)$$

where $\boldsymbol{Z}^{(k)}$ denotes the cluster centers of assign indicator matrix $\boldsymbol{F}^{(k)}$. That problem can be solved via iteratively optimize $\boldsymbol{\zeta}^{(k)}$ and $\boldsymbol{Z}^{(k)}$. Due to space limits, please kindly refer to [Bai and Liang, 2020; Liu and et al, 2023] for more details on the optimization process. After we obtain the optimal solution of $\boldsymbol{\zeta}^{(k)}$, we can further calculate the results of the corresponding cluster distribution $\mathbb{P}(\boldsymbol{C}_j^{(k)}) = \mathcal{N}(\widetilde{\boldsymbol{\mu}}_j^{(k)}, (\widetilde{\boldsymbol{\sigma}}_j^{(k)})^2)$ as:

$$\widetilde{\boldsymbol{\mu}}_j^{(k)} = \frac{\sum_{i=1}^{N} \zeta_{ij}^{(k)} \boldsymbol{\mu}_j^{U_{(k)}}}{\sum_{i=1}^{N} \zeta_{ij}^{(k)}}, \quad \widetilde{\boldsymbol{\sigma}}_j^{(k)} = \frac{\sum_{i=1}^{N} \zeta_{ij}^{(k)} \boldsymbol{\sigma}_j^{U_{(k)}}}{\sum_{i=1}^{N} \zeta_{ij}^{(k)}}.$$

After we obtain these results, we tend to narrow the distance between the local user distribution and these cluster distributions. Therefore, we further provide the compactness co-clustering loss as follows:

$$\ell_C^{(k)} = \sum_{i=1}^{N} \sum_{j=1}^{M} \zeta_{ij}^{(k)} \cdot \mathcal{D}(\mathbb{P}(\boldsymbol{U}_i^{(k)}), \mathbb{P}(\boldsymbol{C}_j^{(k)})). \quad (11)$$

Datasets	Users	Items	Ratings	Density
Douban Movie	800	154,886	93,074	0.075%
Douban Book	800	165,461	29,781	0.022%
Douban Music	800	166,447	30,487	0.023%
Amazon Phone	16,337	9,481	148,271	0.096 %
Amazon Electronics	16,337	40,460	821,301	0.124 %
Amazon Sport	7,857	12,655	163,291	0.164 %
Amazon Cloth	7,857	17,943	187,880	0.133 %
Amazon Game	1,730	12,319	25,036	0.117 %
Amazon Video	1,730	8,751	16,091	0.106 %

Table 1: The statistics information of Douban and Amazon datasets.

Using the compactness co-clustering, the users with similar tastes or preferences will have more compact representations as shown in Fig.1(e). Finally, we can combine the rating prediction loss $\ell_R^{(k)}$, local-global regularization loss $\ell_P^{(k)}$, and compactness co-clustering loss $\ell_C^{(k)}$ for the total loss of proposed FPPDM++:

$$\min \ell_{\texttt{FPPDM++}} = \ell_R^{(k)} + \lambda_P \ell_P^{(k)} + \lambda_C \ell_C^{(k)}, \quad (12)$$

where λ_P and λ_C denote the balance hyper parameters. Note that the FPPDM and FPPDM++ have the same global server aggregation component during training process. The training algorithm details have been provided in Algo.1.

4 Experiments

4.1 Experimental Setup

Datasets. We conduct experiments on two popularly used real-world datasets, i.e., *Douban* and *Amazon*. First, the **Douban** dataset [Zhu *et al.*, 2019; Zhu and et al, 2021] has three domains, i.e., Book, Music, and Movie. Second, the **Amazon** dataset [Zhao *et al.*, 2020; Ni *et al.*, 2019] has six domains, i.e., Phone, Electronics (Elec), Cloth, Sport, Game, and Video. The detailed statistics of these datasets are given in Table 1. For each dataset, we binarize the ratings higher or equal to 4 as positive. We also filter the users and items with less than 5 interactions, following existing research [Zhu *et al.*, 2019; Liu and et al, 2022]. We conduct three cross-domain recommendation tasks in **Amazon** when only two domains (clients) involve ($K = 2$), i.e., **Phone and Elec**, **Cloth and Sport**, and **Game and Video**. Meanwhile we conduct multi-domain recommendation on the whole **Douban** datasets ($K = 3$). We assume that users in different domains are overlapped in the same task, and there are two/three clients involves in **Amazon/Douban** tasks respectively.

Baseline. We compare our proposed **FPPDM** with the following state-of-the-art models. (1) **NeuMF** [He *et al.*, 2017] first adopts deep neural network for collaborative filtering in single domain. (2) **PMLAM** [Ma and et al, 2020] first adopts probabilistic user/item modelling with metric-based loss in single domain. (3) **LightGCN** [He *et al.*, 2020] adopts the graph neural network to model user/item interactions in single domain. (4) **FedMF** [Chai *et al.*, 2020] is the first attempt to adopt federal learning framework across domains. (5) **FedGNN** [Wu *et al.*, 2021] further adopts the graph neural network into federated recommendation. (6) **PriCDR** [Chen and et al, 2022] is the single-target CDR model while sharing the information from the rich domain to the sparse domain.

	(Amazon) Phone & Elec on Phone				(Amazon) Phone & Elec on Elec				(Amazon) Cloth & Sport on Cloth			
	HR@5	NDCG@5	HR@10	NDCG@10	HR@5	NDCG@5	HR@10	NDCG@10	HR@5	NDCG@5	HR@10	NDCG@10
NeuMF	8.28	5.21	13.63	6.89	11.02	6.91	17.85	8.70	5.74	2.66	9.57	5.12
PMLAM	9.65	6.06	15.33	7.97	12.10	7.89	18.44	9.56	6.82	3.39	10.38	5.56
LightGCN	10.23	6.88	16.06	8.55	12.67	8.85	19.30	9.94	7.29	3.81	11.05	6.01
FedMF	9.34	5.93	14.78	8.02	11.91	7.60	17.97	9.33	6.63	3.07	10.22	5.49
FedGNN	10.10	6.76	16.97	8.61	13.28	8.63	19.05	10.12	7.02	3.84	11.26	6.32
PriCDR	11.40	7.61	18.21	9.84	14.05	9.82	20.37	11.24	7.99	5.12	11.67	6.50
DOML	10.59	7.03	17.43	9.15	13.46	8.87	19.62	10.58	7.24	4.56	10.83	6.04
ETL	11.07	7.34	17.95	9.50	13.73	9.43	20.14	11.05	7.87	4.92	11.38	6.26
CDRIB	12.35	8.19	18.87	10.29	14.64	10.04	21.21	12.15	8.30	5.55	12.19	6.81
MSCDR	12.83	8.55	19.10	10.42	14.54	10.23	21.39	11.97	8.58	5.76	11.81	6.93
FPPDM	14.41	9.45	20.82	11.03	15.75	11.18	22.72	12.80	9.47	6.80	13.36	8.13
FPPDM++	**15.02**	**9.96**	**21.44**	**11.50**	**16.31**	**11.37**	**23.53**	**13.26**	**10.05**	**7.38**	**14.14**	**8.42**
	(Amazon) Game & Video on Game				(Amazon) Game & Video on Video				(Amazon) Cloth & Sport on Sport			
	HR@5	NDCG@5	HR@10	NDCG@10	HR@5	NDCG@5	HR@10	NDCG@10	HR@5	NDCG@5	HR@10	NDCG@10
NeuMF	3.25	2.16	6.33	2.64	5.47	2.92	10.54	4.09	5.86	2.93	9.30	4.18
PMLAM	4.07	2.74	6.89	3.17	6.01	3.48	11.30	4.65	6.23	3.32	9.91	4.73
LightGCN	4.62	3.29	7.08	3.45	6.54	3.77	11.86	5.32	6.72	3.71	10.77	5.17
FedMF	3.84	2.52	6.96	2.98	5.70	3.13	11.09	4.81	6.35	3.23	9.84	4.41
FedGNN	4.40	3.13	7.64	3.82	6.59	3.68	11.95	5.47	6.84	3.69	10.08	4.75
PriCDR	5.13	3.38	8.50	4.49	7.22	4.25	12.71	6.28	7.73	4.54	11.20	5.59
DOML	4.56	3.05	7.89	4.01	6.73	3.97	12.24	6.03	7.26	4.11	10.65	5.06
ETL	4.91	3.20	8.17	4.23	7.04	4.32	12.43	5.95	7.55	4.46	11.48	5.80
CDRIB	5.44	3.61	8.51	4.58	7.41	4.65	13.17	6.49	7.90	4.89	12.04	6.22
MSCDR	5.89	3.67	8.78	4.70	7.68	4.80	13.54	6.76	8.13	5.05	11.87	5.97
FPPDM	6.65	4.73	10.05	5.68	9.19	5.76	14.22	7.54	9.31	5.80	12.95	6.78
FPPDM++	**7.53**	**5.20**	**10.87**	**6.05**	**9.92**	**6.44**	**14.76**	**8.11**	**10.02**	**6.58**	**13.60**	**7.49**
	(Douban) Movie				(Douban) Book				(Douban) Music			
	HR@5	NDCG@5	HR@10	NDCG@10	HR@5	NDCG@5	HR@10	NDCG@10	HR@5	NDCG@5	HR@10	NDCG@10
NeuMF	25.36	16.42	45.52	25.23	17.57	10.30	36.43	24.55	26.31	14.58	37.97	22.52
PMLAM	25.80	16.96	46.07	25.76	18.13	10.78	36.95	25.01	26.94	15.29	38.62	23.23
LightGCN	26.47	17.51	46.49	26.68	18.75	11.24	37.46	25.68	27.32	15.73	38.94	23.65
FedMF	26.18	17.10	45.84	26.29	18.02	10.63	36.71	24.82	26.73	14.96	38.35	23.17
FedGNN	26.52	17.43	46.21	26.74	18.60	11.06	37.38	25.14	27.29	15.15	38.70	23.82
PriCDR	27.29	18.23	47.18	27.37	19.36	12.14	38.52	25.90	28.56	16.48	39.53	25.05
DOML	26.84	17.78	46.60	27.12	19.19	11.45	37.96	25.51	27.67	15.63	39.12	24.39
ETL	27.01	18.05	46.94	27.40	19.42	11.91	38.39	25.98	28.10	16.05	39.55	24.71
CDRIB	27.60	18.31	47.37	27.65	19.79	12.53	38.94	26.38	28.82	16.70	40.06	25.24
MSCDR	27.97	18.54	48.09	28.11	20.35	12.80	39.28	26.56	28.75	16.64	40.19	25.40
FPPDM	29.15	19.64	48.86	28.73	21.17	13.58	40.20	27.39	29.43	17.21	40.78	26.02
FPPDM++	**29.79**	**20.07**	**49.32**	**29.28**	**21.56**	**13.91**	**40.75**	**27.65**	**30.10**	**17.53**	**41.24**	**26.48**

Table 2: Experimental results (%) on Douban and Amazon datasets.

	HR@5	NDCG@5	HR@10	NDCG@10
ETL	24.84	15.40	41.63	26.03
ETL-F	25.79	15.89	42.46	26.51
ETL-FC	26.03	16.22	42.97	26.82

Table 3: Method extension results (%) on Douban datasets.

	HR@5	NDCG@5	HR@10	NDCG@10
$M = 5$	26.64	16.85	43.39	27.44
$M = 10$	26.83	17.02	43.56	27.62
$M = 15$	27.15	17.17	43.77	27.80
$M = 20$	27.20	17.19	43.82	27.86

Table 4: The results (%) of tuning M on Douban datasets.

(7) **DOML** [Li and Tuzhilin, 2021] adopts the dual deep transfer module with orthogonal constraints across domains. (8) **ETL** [Chen *et al.*, 2020] utilizes equivalent transformation with autoencoder framework for cross-domain modelling. (9) **CDRIB** [Cao *et al.*, 2022] is the state-of-the-art graph-based cross domain model with information bottleneck theorem. (10) **MSCDR** [Zhao *et al.*, 2022] is the state-of-the-art graph-based multi domain model with attention mechanism. We adopt the same user-item ratings without other auxiliary information for our propose methods and these baselines.

Implemented details. We set batch size $N = 256$ and embedding dimension as $D = 128$ across different domains. We set the hyper parameter $\lambda_P = 0.05$ for local client modelling component in both FPPDM and FPPDM++. The number of graph encoder layer is set to 3 in each clients, following [He *et al.*, 2020; Wang *et al.*, 2019b]. Meanwhile set the hyper parameter $\lambda_C = 0.07$ and the number of clusters as $M = 15$ for compactness co-clustering loss in FPPDM++. Besides, we set the balance hyper parameter $\epsilon = 0.05$ and $\eta = 0.1$. We conduct training for **FedCDR**, FPPDM and FPPDM++ with 5 local epochs per round until converge. We choose Adam [Kingma and Ba, 2014] as optimizer, and adopt HR and NDCG [Wang *et al.*, 2019a] as the evaluation metrics. For all the experiments, we perform five random experiments and report the average results. We report the results measured by the commonly used metrics as $\mathrm{Top@5}$ and $\mathrm{Top@10}$ in Douban and Amazon datasets, respectively.

4.2 Recommendation Performance

The comparison results on Douban and Amazon datasets are shown in Table 2. Note that conventional CDR models (e.g., **DARec**, **DOML**, and **CDRIB**) cannot directly adjust to the multi-domain recommendation in **Douban**. Therefore, we set **Douban Movie** as the source domain and **Douban Book**, **Douban Music** as the target domain for dealing with these CDR models. From that we can conclude: (1) Modelling the user preference distribution can achieve much better results by comparing **PMLAM** with **NeuMF**. However, single domain methods cannot fully utilize the knowledge across domains which make them hard to tackle the sparsity problem. (2) Combining multiple domains information can enhance the model performance by comparing **NeuMF** with **DOML**. Meanwhile, recent models (e.g., **CDRIB**) which adopt graph-based framework even achieve much more promising results than single domain graph-based model (e.g., **LightGCN**). (3) The federal learning framework (e.g., **FedMF**) can even exceed the performance than single domain methods (e.g., **NeuMF**) by combing knowledge and protecting privacy. Nevertheless, the non-deep-learning based **FedMF** cannot handle the complex and complicated user-item relationship. Meanwhile, single-target based **PriCDR** simply transferred knowledge from rich to the sparse domain. Thus it cannot better handle the situation when different domains are both sparse. Latest **FedCDR** model also separates the training and mapping process limits the model performance. (4) Although latent models (e.g., **MetaDPA** and **MSCDR**) reach better performance, they still only extract user/item embeddings which limit their model potentials. These deterministic manner cannot capture and depict complicated user-item relationships. (5) Our proposed FPPDM achieves much satisfied results which indicates the efficacy of FPPDM by integrating preference distribution across different domains. Furthermore, FPPDM++ via compactness co-clustering can gather users with similar characteristics which improve the model performance. Hence FPPDM++ even achieve much better results than proposed FPPDM.

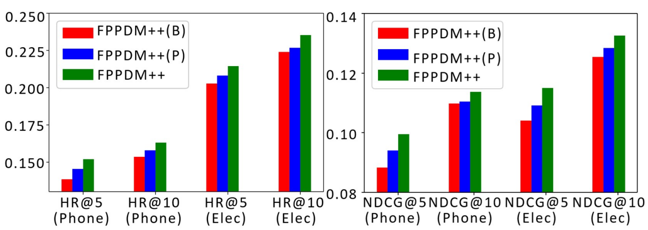

Figure 3: The ablation results (HR and NDCG) on FPPDM++.

4.3 Analysis

Ablation. We compare FPPDM/FPPDM++ with its several variants, including FPPDM++(B) and FPPDM++(P) to study how does each component of FPPDM++ contribute to the final performance. FPPDM++(B) and FPPDM++(P) both adopt the deterministic training manners without estimating the user covariance. While FPPDM++(B) excludes compactness co-clustering module and set $\lambda_C = 0$. We conduct the FPPDM++(B) and FPPDM++(P) on **Amazon Phone and Elec** and report the results (HR and NDCG) in Fig. 3. By comparing the FPPDM and FPPDM++(B) we can conclude that modelling user preference distribution can better capture diverse and complex relationships. However, they cannot gather similar users in the latent space limits the models' accuracy. We can further observe that FPPDM++(P) exceeds the FPPDM++(B) which indicates that clustering users similar tastes or characteristics will boost the model potentials. Moreover, combing distribution modelling and compactness clustering can greatest improve the model performance.

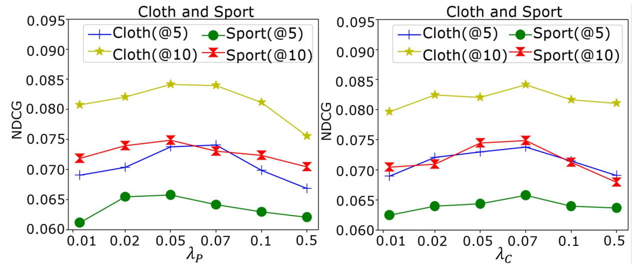

Figure 4: The effect of λ_P and λ_C on FPPDM++.

Method Extension. We further analyse the general extension of our proposed federated framework. To validate this, we apply our proposed FPPDM and FPPDM++ into the MLP-based dual autoencoder model **ETL** as **ETL-F** and **ETL-FC** on conducting the experiments on **Douban** datasets respectively. We report the results HR and NDCG on Table 3. From that we can conclude our proposed federated framework is model-agnostic. That is it can be applied in both graph-based or MLP-based recommendation approaches. Meanwhile utilizing the federated learning strategy by safely sharing knowledge can even enhance the model performance.

Effect of hyper-parameters. We finally study the effects of hyper-parameters on model performance. We first conduct experiments to study the effects of λ_P by varying them in $\{0.01, 0.02, 0.05, 0.07, 0.1, 0.5\}$ for FPPDM++ on **Amazon Cloth and Sport** and report the results in Fig.4. It is difficult to apply local-global regularization alignment for knowledge sharing on each clients when λ_P is too small. While when λ_P is too large, it will slightly hurdle the rating prediction modelling. Therefore we set $\lambda_P = 0.05$ empirically. We further vary λ_C in $\{0.01, 0.02, 0.05, 0.07, 0.1, 0.5\}$ and report the results in Fig.4. Similarly, the bell-shaped curve of λ_C indicated that too large or too small value of λ_C will not suitable in training the model and we set $\lambda_C = 0.07$. Finally, we conduct the experiments on varying the number of cluster M in $\{5, 10, 15, 20\}$ and report the results on Table 4. When M is smaller, it cannot better depict the users with similar tastes. By comparing $M = 20$ and $M = 15$, the performance improvement is rather marginal by consuming more time and space. Hence we set $M = 15$ empirically.

5 Conclusion

In this paper, we investigate the Privacy-Preserving Multi-Domain Recommendation (PPMDR) problem. To tackle this problem, we first propose Probabilistic Preference Distribution Modelling (FPPDM), which includes the *local domain modelling component* and *global server aggregation component*. FPPDM can model and share user/item preference distribution across different domains with federated learning strategy. To better exploit useful semantic information by clustering users with similar characteristics, we further propose FPPDM++ with compactness co-clustering method. We also conduct extensive experiments to demonstrate the superior performance of our proposed FPPDM models.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (No.62172362) and CCF-AFSG Research Fund.

References

[Acar *et al.*, 2021] Durmus Alp Emre Acar, Yue Zhao, Ramon Matas Navarro, Matthew Mattina, Paul N Whatmough, and Venkatesh Saligrama. Federated learning based on dynamic regularization. *arXiv preprint arXiv:2111.04263*, 2021.

[Aono *et al.*, 2017] Yoshinori Aono, Takuya Hayashi, Lihua Wang, Shiho Moriai, et al. Privacy-preserving deep learning via additively homomorphic encryption. *IEEE Transactions on Information Forensics and Security*, 13(5):1333–1345, 2017.

[Bai and Liang, 2020] Liang Bai and Jiye Liang. Sparse subspace clustering with entropy-norm. In *ICML*, pages 561–568. PMLR, 2020.

[Cao *et al.*, 2022] Jiangxia Cao, Jiawei Sheng, Xin Cong, Tingwen Liu, and Bin Wang. Cross-domain recommendation to cold-start users via variational information bottleneck. In *ICDE*, 2022.

[Chai *et al.*, 2020] Di Chai, Leye Wang, Kai Chen, and Qiang Yang. Secure federated matrix factorization. *IEEE Intelligent Systems*, 36(5):11–20, 2020.

[Chen and et al, 2022] Chaochao Chen and et al. Differential private knowledge transfer for privacy-preserving cross-domain recommendation. In *WWW*, pages 1455–1465, 2022.

[Chen *et al.*, 2020] Xu Chen, Ya Zhang, Ivor Tsang, Yuangang Pan, and Jingchao Su. Towards equivalent transformation of user preferences in cross domain recommendation. *arXiv preprint arXiv:2009.06884*, 2020.

[Dong *et al.*, 2022] Xin Dong, Sai Qian Zhang, Ang Li, and HT Kung. Spherefed: Hyperspherical federated learning. In *European Conference on Computer Vision*, pages 165–184. Springer, 2022.

[Elkahky *et al.*, 2015] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *WWW*, pages 278–288, 2015.

[Fan, 1949] Ky Fan. On a theorem of weyl concerning eigenvalues of linear transformations i. *Proceedings of the National Academy of Sciences*, 35(11):652–655, 1949.

[Gentry, 2009] Craig Gentry. *A fully homomorphic encryption scheme*. Stanford university, 2009.

[Guo *et al.*, 2021] Lei Guo, Li Tang, Tong Chen, Lei Zhu, Quoc Viet Hung Nguyen, and Hongzhi Yin. Da-gcn: a domain-aware attentive graph convolution network for shared-account cross-domain sequential recommendation. *arXiv preprint arXiv:2105.03300*, 2021.

[He *et al.*, 2017] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pages 173–182, 2017.

[He *et al.*, 2020] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, pages 639–648, 2020.

[Hu *et al.*, 2018] Guangneng Hu, Yu Zhang, and Qiang Yang. Conet: Collaborative cross networks for cross-domain recommendation. In *CIKM*, pages 667–676, 2018.

[Jiang *et al.*, 2020] Junyang Jiang, Deqing Yang, Yanghua Xiao, and Chenlu Shen. Convolutional gaussian embeddings for personalized recommendation with uncertainty. *arXiv preprint arXiv:2006.10932*, 2020.

[Kingma and Ba, 2014] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *Computer Science*, 2014.

[Li and Tuzhilin, 2021] Pan Li and Alexander Tuzhilin. Dual metric learning for effective and efficient cross-domain recommendations. *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[Li *et al.*, 2022] Zhi Li, Daichi Amagata, Yihong Zhang, Takahiro Hara, Shuichiro Haruta, Kei Yonekawa, and Mori Kurokawa. Debiasing graph transfer learning via item semantic clustering for cross-domain recommendations. *ArXiv*, abs/2211.03390, 2022.

[Lian *et al.*, 2017] Jianxun Lian, Fuzheng Zhang, Xing Xie, and Guangzhong Sun. Cccfnet: a content-boosted collaborative filtering neural network for cross domain recommender systems. In *WWW*, pages 817–818, 2017.

[Lin *et al.*, 2020] Tao Lin, Lingjing Kong, Sebastian U Stich, and Martin Jaggi. Ensemble distillation for robust model fusion in federated learning. *Advances in Neural Information Processing Systems*, 33:2351–2363, 2020.

[Lin *et al.*, 2022a] Shuai Lin, Chen Liu, Pan Zhou, Zi-Yuan Hu, Shuojia Wang, Ruihui Zhao, Yefeng Zheng, Liang Lin, Eric Xing, and Xiaodan Liang. Prototypical graph contrastive learning. *TNNLS*, 2022.

[Lin *et al.*, 2022b] Zihan Lin, Changxin Tian, Yupeng Hou, and Wayne Xin Zhao. Improving graph collaborative filtering with neighborhood-enriched contrastive learning. In *WWW*, pages 2320–2329, 2022.

[Liu and et al, 2022] Weiming Liu and et al. Exploiting variational domain-invariant user embedding for partially overlapped cross domain recommendation. In *SIGIR*, pages 312–321, 2022.

[Liu and et al, 2023] Weiming Liu and et al. Contrastive proxy kernel stein path alignment for cross-domain cold-start recommendation. *TKDE*, 2023.

[Liu *et al.*, 2020] Meng Liu, Jianjun Li, Guohui Li, and Peng Pan. Cross domain recommendation via bi-directional transfer graph collaborative filtering networks. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 885–894, 2020.

[Ma and et al, 2020] Chen Ma and et al. Probabilistic metric learning with adaptive margin for top-k recommendation. In *KDD*, pages 1036–1044, 2020.

[McMahan *et al.*, 2017] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

[Meihan *et al.*, 2022] Wu Meihan, Li Li, Chang Tao, Eric Rigall, Wang Xiaodong, and Xu Cheng-Zhong. Fedcdr: Federated cross-domain recommendation for privacy-preserving rating prediction. In *CIKM*, pages 2179–2188, 2022.

[Ni *et al.*, 2019] Jianmo Ni, Jiacheng Li, and Julian McAuley. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *EMNLP-IJCNLP*, pages 188–197, 2019.

[Nie *et al.*, 2014] Feiping Nie, Xiaoqian Wang, and Heng Huang. Clustering and projected clustering with adaptive neighbors. In *KDD*, pages 977–986, 2014.

[Pillutla *et al.*, 2022] Krishna Pillutla, Kshitiz Malik, Abdel-Rahman Mohamed, Mike Rabbat, Maziar Sanjabi, and Lin Xiao. Federated learning with partial model personalization. In *ICML*, pages 17716–17758. PMLR, 2022.

[Sun and et al, 2022] Zhu Sun and et al. Daisyrec 2.0: Benchmarking recommendation for rigorous evaluation. *TPAMI*, 2022.

[T Dinh *et al.*, 2020] Canh T Dinh, Nguyen Tran, and Josh Nguyen. Personalized federated learning with moreau envelopes. *NeurIPS*, 33:21394–21405, 2020.

[Tan and et al, 2021] Yue Tan and et al. Fedproto: Federated prototype learning over heterogeneous devices. *AAAI*, 2021.

[Wang and et al, 2022] Hao Wang and et al. Escm2: Entire space counterfactual multi-task model for post-click conversion rate estimation. In *SIGIR*, pages 363–372, 2022.

[Wang *et al.*, 2019a] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 950–958, 2019.

[Wang *et al.*, 2019b] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. Neural graph collaborative filtering. In *SIGIR*, pages 165–174, 2019.

[Wu *et al.*, 2021] Chuhan Wu, Fangzhao Wu, Yang Cao, Yongfeng Huang, and Xing Xie. Fedgnn: Federated graph neural network for privacy-preserving recommendation. *arXiv preprint arXiv:2102.04925*, 2021.

[Yuan *et al.*, 2019] Feng Yuan, Lina Yao, and Boualem Benatallah. Darec: Deep domain adaptation for cross-domain recommendation via transferring rating patterns. *IJCAI*, 2019.

[Zang *et al.*, 2021a] Tianzi Zang, Yanmin Zhu, Haobing Liu, Ruohan Zhang, and Jiadi Yu. A survey on cross-domain recommendation: Taxonomies, methods, and future directions. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 2021.

[Zang *et al.*, 2021b] Tianzi Zang, Yanmin Zhu, Haobing Liu, Ruohan Zhang, and Jiadi Yu. A survey on cross-domain recommendation: taxonomies, methods, and future directions. *arXiv preprint arXiv:2108.03357*, 2021.

[Zhang *et al.*, 2021] Xinwei Zhang, Mingyi Hong, Sairaj Dhople, Wotao Yin, and Yang Liu. Fedpd: A federated learning framework with adaptivity to non-iid data. *IEEE Transactions on Signal Processing*, 69:6055–6070, 2021.

[Zhang *et al.*, 2022a] Xu Zhang, Yinchuan Li, Wenpeng Li, Kaiyang Guo, and Yunfeng Shao. Personalized federated learning via variational bayesian inference. In *ICML*, pages 26293–26310. PMLR, 2022.

[Zhang *et al.*, 2022b] Yan Zhang, Changyu Li, Ivor W Tsang, Hui Xu, Lixin Duan, Hongzhi Yin, Wen Li, and Jie Shao. Diverse preference augmentation with multiple domains for cold-start recommendations. *ICDE*, 2022.

[Zhao *et al.*, 2019] Cheng Zhao, Chenliang Li, and Cong Fu. Cross-domain recommendation via preference propagation graphnet. In *CIKM*, pages 2165–2168, 2019.

[Zhao *et al.*, 2020] Cheng Zhao, Chenliang Li, Rong Xiao, Hongbo Deng, and Aixin Sun. Catn: Cross-domain recommendation for cold-start users via aspect transfer network. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 229–238, 2020.

[Zhao *et al.*, 2022] Xiaoyun Zhao, Ning Yang, and Philip S Yu. Multi-sparse-domain collaborative recommendation via enhanced comprehensive aspect preference learning. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 1452–1460, 2022.

[Zheng *et al.*, 2022] Xiaolin Zheng, Jiajie Su, Weiming Liu, and Chaochao Chen. Ddghm: Dual dynamic graph with hybrid metric training for cross-domain sequential recommendation. *Proceedings of the 30th ACM International Conference on Multimedia*, 2022.

[Zhu and et al, 2021] Feng Zhu and et al. A unified framework for cross-domain and cross-system recommendations. *TKDE*, 2021.

[Zhu *et al.*, 2019] Feng Zhu, Chaochao Chen, Yan Wang, Guanfeng Liu, and Xiaolin Zheng. Dtcdr: A framework for dual-target cross-domain recommendation. In *CIKM*, pages 1533–1542, 2019.

[Zhu *et al.*, 2021] Feng Zhu, Yan Wang, Chaochao Chen, Jun Zhou, Longfei Li, and Guanfeng Liu. Cross-domain recommendation: Challenges, progress, and prospects. *arXiv preprint arXiv:2103.01696*, 2021.