Supplementary Materials: Notation, Reproducibility, Target Problems, and Experiments

Anonymous Authors

For a better understanding of this paper, in this section, we provide more details of notations, reproducibility, target problems, and experimental results.

1 IMPORTANT NOTATIONS

For the sake of better readability, we list the main notations of this paper in Table 1.

Table 1: Important notations

Symbol	Definition
D^x	Domain x, $x \in \{0, 1, .., a-1\}$, a is the number of domains
$R \in \mathbb{R}^{m \times n}$	the rating matrix
$r_{ij} \in R$	the rating of user u_i on item v_j
k	the dimension of user/item embeddings
m	the number of users
n	the number of items
$\mathcal{U} = \{u_1, ..., u_m\}$	the set of users
$U \in \mathbb{R}^{m \times k}$	the embedding matrix of users
$\mathcal{V} = \{v_1, ..., v_n\}$	the set of items
$V \in \mathbb{R}^{n \times k}$	the embedding matrix of items
$G = (\{\mathcal{U}, \mathcal{V}\}, E)$	the heterogeneous graph, E is the set of user-item interaction relationships (from R)
$\tilde{U}/\tilde{V}$	the combined embeddings of common users/items
$y_{ij} \in Y$	the interaction of user u_i on item v_j
$Y \in \mathbb{R}^{m \times n}$	the user-item interaction matrix
$*^x, x \in \{0, 1, 2, ..., a-1\}$	the notations for domain x, where a is the total number of domains, e.g., m^0 represents the number of users in domain 0
$\hat{*}$	the predicted notations, e.g., $\hat{y}_{ij}$ represents the predicted interaction of u_i on item v_j

2 REPRODUCIBILITY DETAILS

Apart from the observed user-item interaction samples as positive samples, denoted as Y^+, we tend to randomly select a certain number of unobserved user-item interactions as negative samples, denoted by Y_s^-, to replace the whole negative sample set Y^-, i.e., all unobserved interactions. This is because we cannot observe users' attitudes toward their unrated items. It could be that the users have not been aware of these unrated items because of ranking position or display style. This training strategy has been widely used in the existing approaches [2, 9, 11]. Based on rating information, the label of user-item interaction y_{ij} between a user u_i and an item v_i can be represented as:

$$y_{ij} = \begin{cases} r_{ij}, & \text{if } y_{ij} \in Y^+; \\ 0, & \text{if } y_{ij} \in Y_s^-; \\ null, & \text{otherwise}. \end{cases} \tag{1}$$

Most of the existing CDR approaches [2, 7, 11], and our AMA-CDR choose a normalised cross-entropy loss as follows.

$$\ell(y, \hat{y}) = -\frac{y}{max(R)} \log \hat{y} - (1 - \frac{y}{max(R)}) \log(1 - \hat{y}), \tag{2}$$

where $max(R)$ is the maximum rating in a domain.

3 SUPPLEMENTARY DETAILS OF THE TARGET PROBLEMS

3.1 Many-to-Many CDR

We define our main task, i.e., many-to-many cross-domain recommendation, as follows. **Many-to-many cross-domain recommendation**: Given the multiple (many) domains 0 to $a-1$, including user sets $\{\mathcal{U}^0, ..., \mathcal{U}^{a-1}\}$ and item sets $\{\mathcal{V}^0, ..., \mathcal{V}^{a-1}\}$, many-to-many CDR is to improve the recommendation accuracy in all domains simultaneously by leveraging their observed information.

3.2 Transfer Paradigms in CDR

In the area of cross-domain recommendation, as introduced in [10], the existing CDR approaches can be generally classified into three groups, i.e., single-target, dual-target, and multi-target. However, as more and more CDR methods emerge, according to their transfer paradigms, the existing CDR approaches can be further divided into four groups, i.e., single-target: one-to-one paradigm and many-to-one paradigm, and dual/multi-target: one-to-many paradigm and many-to-many paradigm, which serve different cross-domain scenarios. For example, the many-to-many paradigm represents that the CDR approach leverages the knowledge learned from many (multiple) domains to improve the recommendation accuracy in all domains simultaneously. Compared with the first three paradigms, the many-to-many paradigm is more challenging and meaningful. This is because the many-to-many paradigm can almost serve all CDR scenarios.

3.3 Objective Distortion

Many existing CDR approaches ignore the importance of their training paradigms, i.e., two-step training and end-to-end training. Some embedding-based transfer CDR approaches, e.g., EMCDR [3], DCDCSR [8], GA-MTCDR [11] and PCRec [5], tend to first train the embeddings of users/items (step 1) by some embedding generation models, e.g., matrix factorization [3, 8] or graph embedding [5, 9, 11], and then transfer/share the embeddings across domains (step 2). This is a typical solution of two-step training, which may lead to objective distortion between the two steps. The two-step training paradigms of these CDR approaches can be generally classified into two strategies (see the two strategies in Eq. (3)) and there are respectively represented as follows.

$$\text{Step 1:}\begin{cases}\underbrace{\min_{U^{1,x},V^{1,x},\Theta^{1,x}}\sum_{y\in Y^{x+}\cup Y_s^{x-}}\ell(\hat{y},y),\ \hat{y}=\sigma(f(U_i^{1,x},V_j^{1,x},\Theta^{1,x})),}_{\text{Strategy 1}}\\ \underbrace{\min_{U^{1,x},V^{1,x},\Theta^{1,x}} obj(U^{1,x},V^{1,x},\Theta^{1,x}),}_{\text{Strategy 2}}\end{cases}\tag{3}$$

where $\ell(*)$ is the loss function, $\hat{y}$ is the interaction prediction between the user u_i and the item v_j, $\sigma(*)$ is the activation function, $f(*)$ is the prediction function, e.g., matrix factorization(MF) and multilayer perception (MLP), $U_i^{1,x}$ is the pre-trained user embedding of u_i in the domain D^x in step 1, $V_j^{1,x}$ is the pre-trained item embedding of v_i in the domain D^x in step 1, Θ_1^x are the parameters for the prediction function $f(*)$ in step 1, and $obj(*)$ is an independent objective function, e.g., the neighbourhood preserving objective of graph embedding models (such as DeepWalk [4] and Node2vec [1]).

$$\text{Step 2:}\begin{cases}\min_{U^{2,x},V^{2,x},\Theta^{2,x}}\sum_{y\in Y^{x+}\cup Y_s^{x-}}\ell(\hat{y},y),\ \hat{y}=\sigma(f(U^{2,x},V^{2,x},\Theta^{2,x})),\\ U^{2,x}=t(U^{1,0},...,U^{1,a-1}),\ V^{2,x}=t(V^{1,0},...,V^{1,a-1}),\end{cases}\tag{4}$$

where $U^{2,x}$ is the optimized user embedding set generated by the transfer function $t(*)$, e.g., non-linear mapping [3, 8] and embedding combination [9], in step 2. Similarly, $V^{2,x}$ is the optimized item embedding set generated by transfer strategies. Also, $\Theta^{2,x}$ are the parameters in step 2.

$$\text{End-to-end:}\begin{cases}\min_{U^x,V^x,\Theta^x}\sum_{y\in Y^{x+}\cup Y_s^{x-}}\ell(\hat{y},y),\ \hat{y}=\sigma(f(U^x,V^x,\Theta^x)),\\ U^x=t(U^0,...,U^{a-1}),\ V^x=t(V^0,...,V^{a-1}).\end{cases}\tag{5}$$

Analysis The two steps do not share any parameters, i.e., $\Theta^{1,x}$ and $\Theta^{2,x}$ are definitely different, and thus the training process of step 1 can not serve for step 2. Moreover, as for Strategy 2 in Step 1, there are the independent objectives of embedding generation models, which are totally different from the objectives of Step 2, i.e., recommendation models. These factors lead to a significant objective distortion between the two steps. In contrast to the two-step training, the most outstanding advantage of end-to-end training is to optimize only one objective, i.e., the direct objective for recommendations.

3.4 Negative Transfer

The traditional definition of negative transfer is that transferring knowledge from the source domain can have a negative impact on the target learner [6], which focuses on the two domains, i.e., a source domain and a target domain. However, there are multiple source and target domains in many-to-many CDR scenarios. Therefore, we define the notion of negative transfer (caused by undifferentiated knowledge transfer) in the many-to-many CDR scenarios as follows.

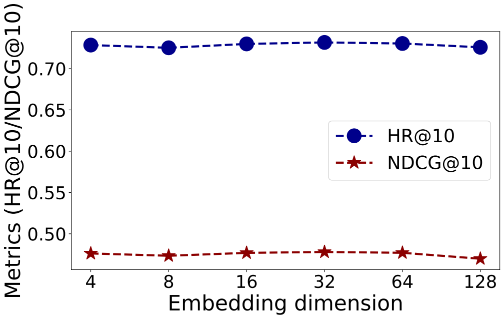

Figure 1: The impact of embedding dimension.

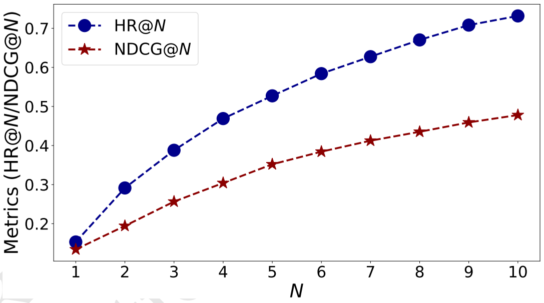

Figure 2: The impact of top-N.

$$\begin{aligned}&\sum_{y\in Y^{x+}\cup Y_s^{x-}}\ell(\hat{y}^{new},y)\ >\sum_{y\in Y^{x+}\cup Y_s^{x-}}\ell(\hat{y}^{pre},y),\\ &\hat{y}^{new}=\sigma(f(t(U^0,...,U^{a-2},U^{a-1}),t(V^0,...,V^{a-2},V^{a-1}),\Theta^x)),\\ &\hat{y}^{pre}=\sigma(f(t(U^0,...,U^{a-2}),t(V^0,...,V^{a-2}),\Theta^x)),\end{aligned}\tag{6}$$

where $\ell(*)$ is the loss function, $\hat{y}$ is the interaction prediction, $\sigma(*)$ is the activation function, $f(*)$ is the prediction function, e.g., matrix factorization(MF) and multilayer perception (MLP), and $t(*)$ is the transfer function, e.g., non-linear mapping [3, 8] and embedding combination [9]. This definition represents that after adding a new source domain, i.e., D^{a-1} with the embeddings of users/items U^{a-1}/V^{a-1}, the training loss of transfer methods (the new one) in the domain D^x is larger than that of the previous one. In fact, in our experiments, we directly choose recommendation performance rather than training loss to judge whether the negative transfer occurs or not.

4 EXPLORATORY EXPERIMENTS

To validate the impacts of the hyper-parameters, i.e., embedding dimension k and top-N, on the performance of our AMA-CDR model, we additionally conduct the following two experiments. Due to space limitations, we only report the experimental results on the DoubanBook domain.

4.1 Impact of embedding dimension

To study the impact of embedding dimension on our proposed AMA-CDR, we choose different dimensions, i.e., {4, 8, 16, 32, 64, 128}, for performance comparison. The experimental results of HR@10 and NDCG@10 on the DoubanBook domain in Figure 1. As observed from Figure 1, when the embedding dimension k is 32, our AMA-CDR achieves the best performance. From 32 to 128, the performance decreases with k mainly because a large dimension k may make the parameters of our AMA-CDR grow exponentially and thus lead to the over-fitting problem.

4.2 Impact of top-N recommendation

To study the impact of top-N recommendation, we compare the recommendation performance of our AMA-CDR in terms of HR@N and NDCG@N, where N ranges from 1 to 10. Since the performance trends of all top-N experiments on all the seven domains, due to space limitations, we only report the results on the DoubanBook domain in Figure 2. As we can see from Figure 2, the recommendation metrics of our AMA-CDR increase with N in the DoubanMusic domain. However, this does not mean that in real applications, we need to choose a large N for recommendations because users cannot really check a long recommendation list. We should achieve a balance between good recommendation performance and good user experience.

REFERENCES

[1] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. 855–864.

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

[3] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-Domain Recommendation: An Embedding and Mapping Approach. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. 2464–2470.

[4] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 701–710.

[5] Chen Wang, Yueqing Liang, Zhiwei Liu, Tao Zhang, and Philip S Yu. 2021. Pre-training Graph Neural Network for Cross Domain Recommendation. *arXiv preprint arXiv:2111.08268* (2021).

[6] Zirui Wang, Zihang Dai, Barnabás Póczos, and Jaime Carbonell. 2019. Characterizing and avoiding negative transfer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 11293–11302.

[7] Hong-Jian Xue, Xinyu Dai, Jianbing Zhang, Shujian Huang, and Jiajun Chen. 2017. Deep Matrix Factorization Models for Recommender Systems. In *Proceedings of the 26th international joint conference on artificial intelligence*. 3203–3209.

[8] Feng Zhu, Yan Wang, Chaochao Chen, Guanfeng Liu, Mehmet A Orgun, and Jia Wu. 2018. A Deep Framework for Cross-Domain and Cross-System Recommendations. In *IJCAI International Joint Conference on Artificial Intelligence*. 3711–3717.

[9] Feng Zhu, Yan Wang, Chaochao Chen, Guanfeng Liu, and Xiaolin Zheng. 2020. A Graphical and Attentional Framework for Dual-Target Cross-Domain Recommendation. In *29th International Joint Conference on Artificial Intelligence*. 3001–3008.

[10] Feng Zhu, Yan Wang, Chaochao Chen, Jun Zhou, Longfei Li, and Guanfeng Liu. 2021. Cross-domain recommendation: challenges, progress, and prospects. (2021).

[11] Feng Zhu, Yan Wang, Jun Zhou, Chaochao Chen, Longfei Li, and Guanfeng Liu. 2021. A unified framework for cross-domain and cross-system recommendations. *IEEE Transactions on Knowledge and Data Engineering* (2021).