Correlative Preference Transfer with Hierarchical Hypergraph Network for Multi-Domain Recommendation

Zixuan Xu‡, Penghui Wei‡, Shaoguo Liu*, Liang Wang and Bo Zheng
Alibaba Group
Beijing, China
{xuzixuan.xzx,wph242967,shaoguo.lsg,liangbo.wl,bozheng}@alibaba-inc.com

WWW 2023

Reported by Nengqiang Xiang

Introduction

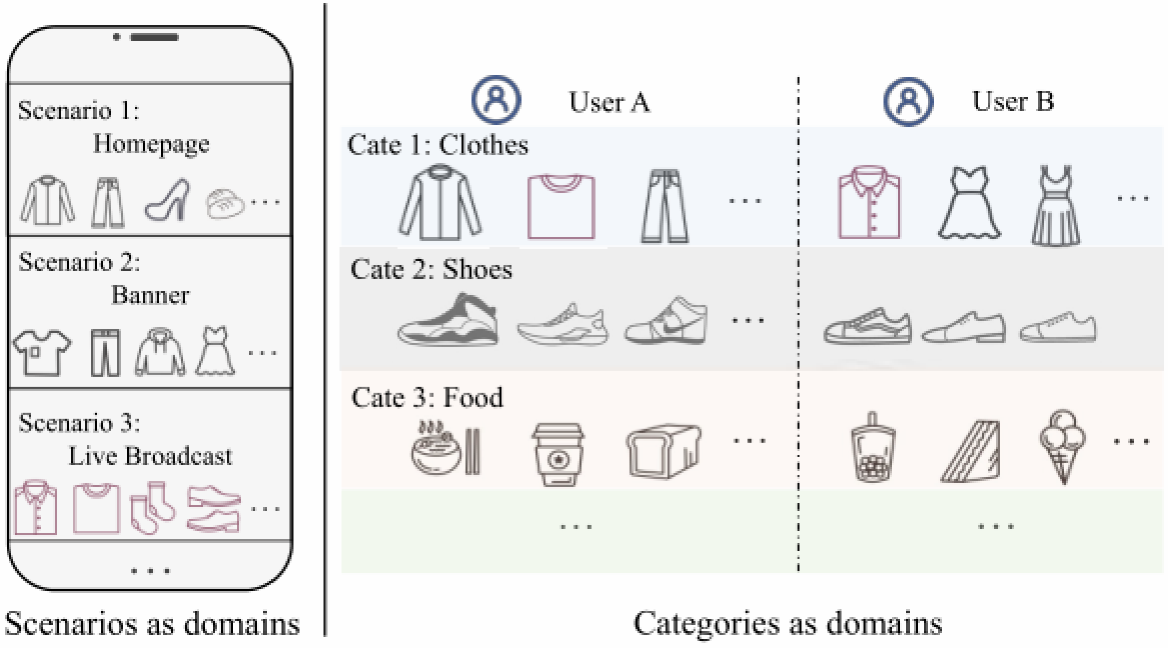


Figure 1: Illustration of multi-domain recommendation. The definition of *domain* can be *recommendation scenario* or *item categories*.

Conventional graph neural network based methods usually deal with each domain separately, or train a shared model to serve all domains. The former fails to leverage users' cross-domain behaviors, making the behavior sparseness issue a great obstacle. The latter learns shared user representation with respect to all domains, which neglects users' domain-specific preferences.

In this paper, the author proposes H3Trans, a hierarchical hypergraph network based correlative preference transfer framework for MDR, which represents multi-domain user-item interactions into a unified graph to help preference transfer.

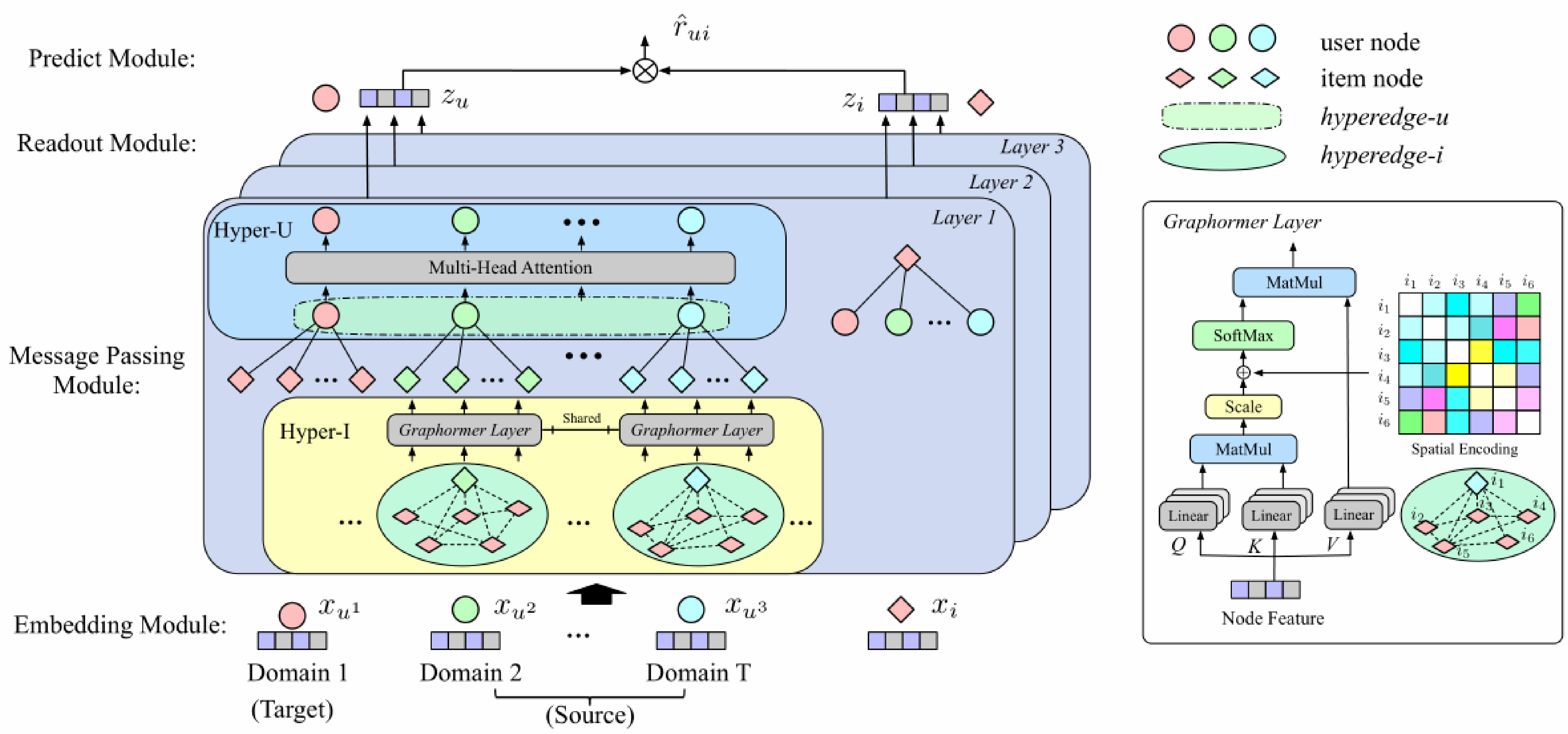


Figure 2: Overall architecture of H^{3}Trans. It contains two hyperedge-based modules: adaptive user aggregation (Hyper-U) and dynamic item transfer module (Hyper-I). These two modules compose a hierarchical hypergraph neural network. Different colors refer to different domains. Here we regard the first domain $\mathcal{D}_1$ as target domain and the others are sources.

PRELIMINARIES:

$\{\mathcal{D}_m\}_{m=1}^{T}$ $\quad U^m \quad I^m$

$$\hat{r}_{u,i}^{m} = f(z_u, z_i \mid \mathcal{D}_m) \tag{1}$$

$\mathcal{R}^m \in \mathbb{R}^{|U^m| \times |I^m|}$

$\{\mathcal{R}^m\}_{m=1}^{T}$

$U = U^1 = U^2 = \cdots = U^T$

$I = I^1 \cup I^2 \cup \cdots \cup I^T$

Unified Multi-domain Graph:

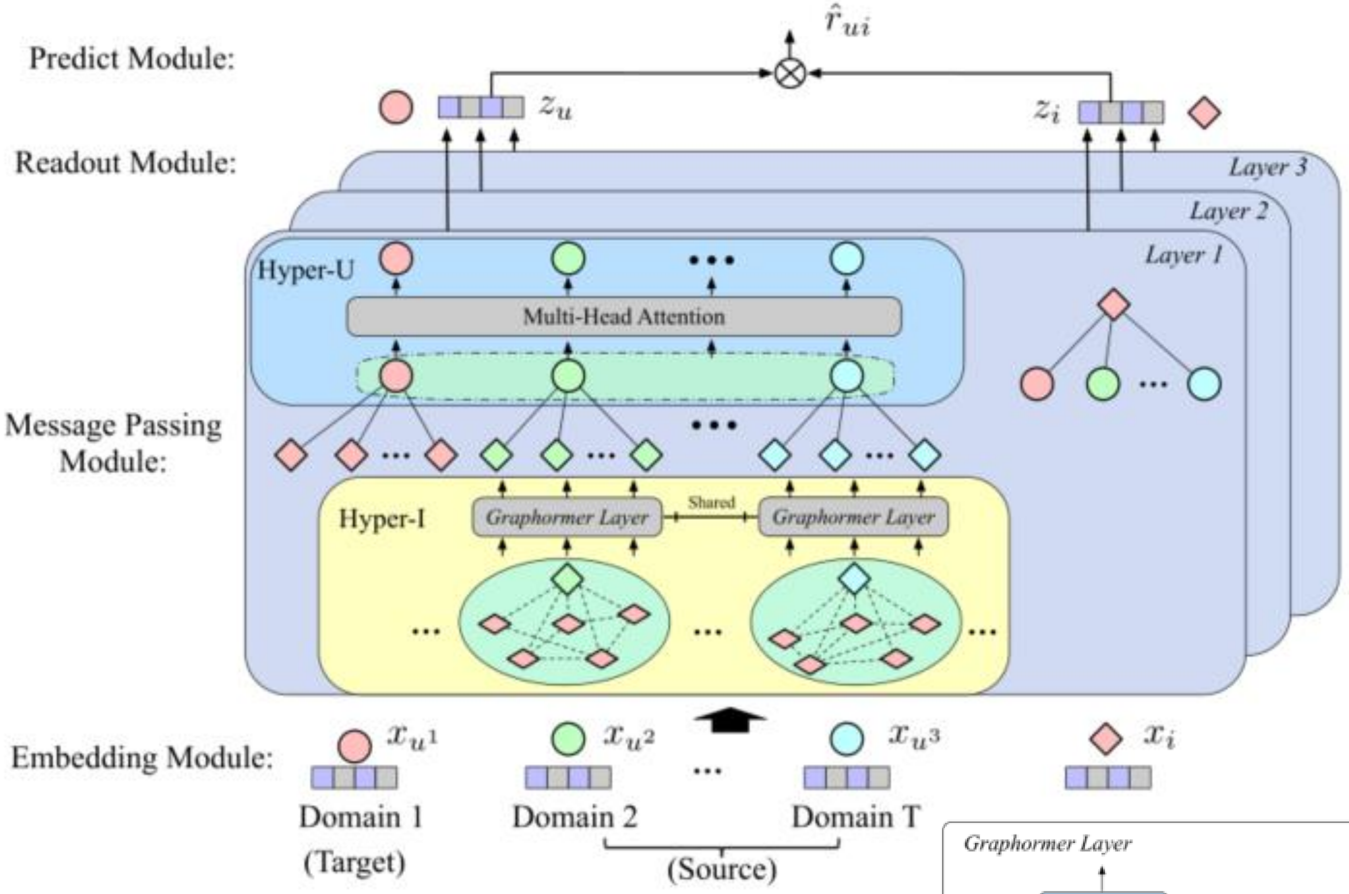
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$ $\quad \mathcal{V} = \mathcal{U} \cup \mathcal{I}$

$u \in U = (u^1, u^2, \cdots, u^T)$

$i \in I$

$\mathcal{E} = \mathcal{E}^1 \cup \mathcal{E}^2 \cup \cdots \cup \mathcal{E}^T$

user set:



Message Passing Module：

Neighbor aggregation:

$$
\begin{aligned}
h^{(l)}_{\mathcal{N}_{u^m}} &= \mathrm{AGG}_{\mathrm{U}}\left(\left\{h_i^{(l-1)} \mid i \in \mathcal{N}_{u^m}\right\}\right)\\
h^{(l)}_{\mathcal{N}_i} &= \mathrm{AGG}_{\mathrm{I}}\left(\left\{h_{u^m}^{(l-1)} \mid u^m \in \mathcal{N}_i\right\}\right)
\end{aligned}
\tag{2}
$$

hyperedge-i

Path-based: $i \rightarrow u^s \rightarrow u^t \rightarrow j$

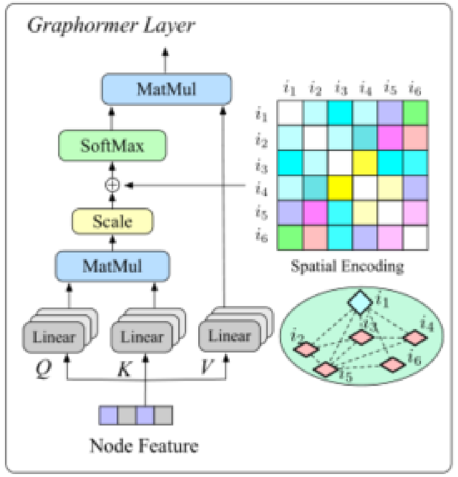
Embedding-based: From target domain find the top-k similar items

Graphormer Layer：

$$
h_i^{(l-1)} \leftarrow \mathrm{GH}_{\mathrm{HyperI}}\left(\mathrm{Concat}\left(h_i^{(l-1)}, \left\{h_j^{(l-1)} \mid |j \in \mathcal{S}_i^t\right\}\right)\right)[0] \tag{6}
$$

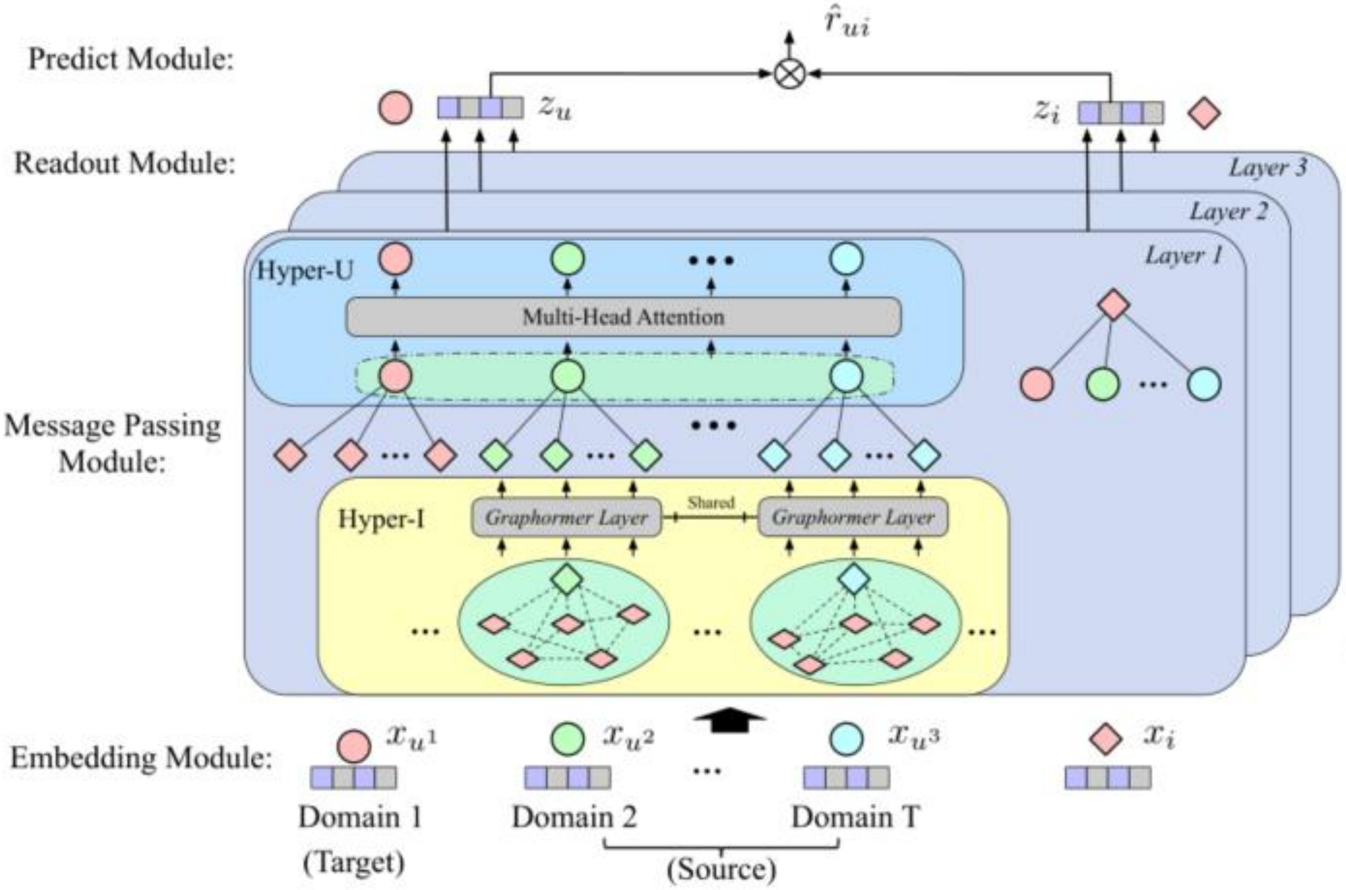
where $\mathrm{GH}_{\mathrm{HyperI}}(\cdot)$ is the graphormer layer for Hyper-I module:

$$
\begin{aligned}
\mathrm{GH}_{\mathrm{HyperI}}(H_{\mathrm{I}}) &= \mathrm{Concat}\left(\mathrm{Attn}_{\mathrm{I},1}(H_{\mathrm{I}}), \cdots, \mathrm{Attn}_{\mathrm{I},P}(H_{\mathrm{I}})\right) W_{\mathrm{I}}^{O},\\
\mathrm{Attn}_{\mathrm{I},p}(H_{\mathrm{I}}) &= \mathrm{softmax}\left(\frac{Q_{\mathrm{I},p} K_{\mathrm{I},p}^{\top}}{\sqrt{d_{h_i}/P}} + \Phi(B)\right) V_{\mathrm{I},p},\\
Q_{\mathrm{I},p} &= H_{\mathrm{I}} W_{\mathrm{I},p}^{Q}, K_{\mathrm{I},p} = H_{\mathrm{I}} W_{\mathrm{I},p}^{K}, V_{\mathrm{I},p} = H_{\mathrm{I}} W_{\mathrm{I},p}^{V}
\end{aligned}
\tag{7}
$$



Embedding Module.：

look-up table $X \in \mathbb{R}^{(|\mathcal{U}|+|\mathcal{I}|)\times d}$



Node update:

$$h_{u^m}^{(l)} = \mathrm{UP_U}\left(h_{u^m}^{(l-1)}, h_{N_{u^m}}^{(l)}\right)$$
$$h_i^{(l)} = \mathrm{UP_I}\left(h_i^{(l-1)}, h_{N_i}^{(l)}\right) \tag{3}$$

hyperedge-u

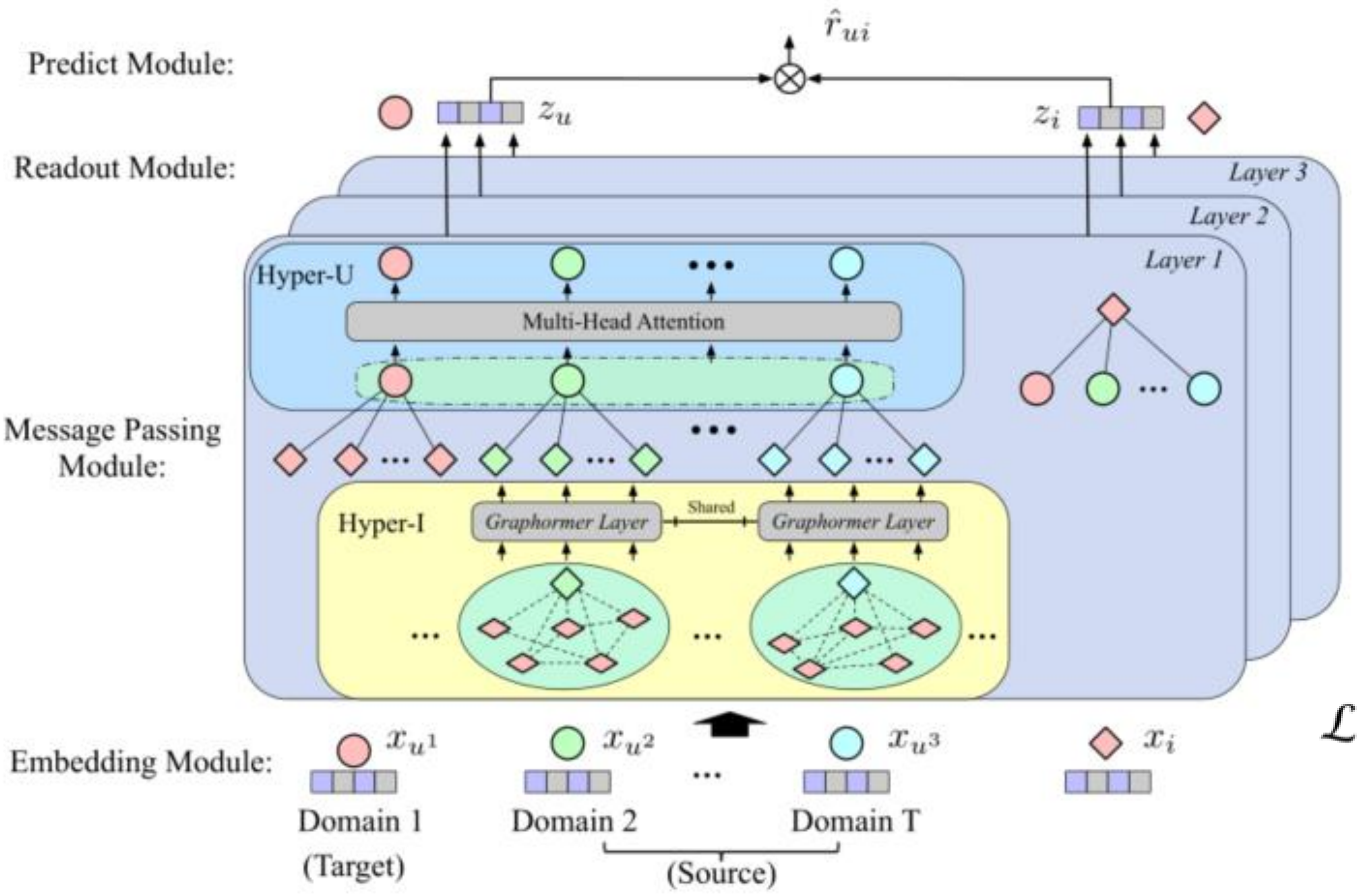
$$u \in \boldsymbol{U} = (u^1, u^2, \cdots, u^T)$$

Multi-head Attention Layer:

$$\left[h_{u^1}^{(l)}, h_{u^2}^{(l)}, \cdots, h_{u^T}^{(l)}\right] \leftarrow \mathrm{MHA_{HyperU}}\left(\left[h_{u^1}^{(l)}, h_{u^2}^{(l)}, \cdots, h_{u^T}^{(l)}\right]\right), \tag{8}$$

where $\mathrm{MHA_{HyperU}}(\cdot)$ denotes the multi-head attention layer:

$$\mathrm{MHA_{HyperU}}(H_U) = \mathrm{Concat}\left(\mathrm{Attn}_{U,1}(H_U), \cdots, \mathrm{Attn}_{I,P}(H_U)\right) W_U^O,$$
$$\mathrm{Attn}_{U,p}(H_U) = \mathrm{softmax}\left(\frac{Q_{U,p} K_{U,p}^{\top}}{\sqrt{d_{h_u}/P}}\right) V_{U,p},$$
$$Q_{U,p} = H_U W_{U,p}^Q, K_{U,p} = H_U W_{U,p}^K, V_{U,p} = H_U W_{U,p}^V \tag{9}$$



Readout Module:

$$z_v = \text{Readout}\left(h_v^{(l)} \mid l \in [1, \ldots, L]\right), \tag{4}$$

Prediction Module:

$$\hat{r}_{u,i}^{m} = f(z_{u^m}, z_i) \tag{5}$$

Model Optimization :

$$\mathcal{L}(u, i \mid \mathcal{D}_m) = -\log \frac{\exp(sim(z_u^m, z_i)/\tau)}{\sum_{i_-} \exp(sim(z_u^m, z_{i_-})/\tau)} \tag{10}$$

(u^m, i_-) is a randomly sampled negative pair that $r_{u,i_-}^{m} = 0$

Experiments

Table 1: Dataset Statistics

Product Dataset				Public Amazon Dataset			
Domains	#user	#item	#click	Domains	#user	#item	#click
MDR-A	84.6M	6.3M	3.1B	Books	1.67M	0.99M	26.8M
MDR-B	34.0M	1.4M	0.6B	Music	0.11M	0.12M	1.5M
MDR-C	24.7M	0.5M	0.3B	Movie	0.23M	0.08M	3.1M
MDR-D	29.1M	0.6M	0.2B	-	-	-	-

Table 2: Main results on product dataset

Method	MDR-A			MDR-B			MDR-C			MDR-D		
	Mrr	HR@20	HR@50	Mrr	HR@20	HR@50	Mrr	HR@20	HR@50	Mrr	HR@20	HR@50
Base	0.0368	2.37%	6.46%	0.0625	4.87%	12.60%	0.0640	4.78%	13.20%	0.0753	5.11%	12.65%
PPGN (Mix)	0.0481	2.98%	8.47%	0.0603	4.22%	11.61%	0.1017	8.58%	18.99%	0.1131	7.60%	17.59%
MGNN	0.0544	3.68%	8.11%	0.0699	5.34%	14.28%	0.1079	12.22%	21.34%	0.1428	10.67%	21.81%
PCRec	0.0635	4.38%	9.71%	0.0845	7.31%	16.63%	0.1546	14.71%	25.99%	0.1738	15.16%	26.59%
BiTGCF	0.0663	4.59%	10.61%	0.0986	8.66%	18.46%	0.1591	15.48%	26.49%	0.1577	13.73%	23.66%
BiTGCF+	0.0750	5.08%	12.31%	0.1237	9.87%	20.71%	0.1713	16.15%	28.63%	0.1685	14.76%	25.85%
H^3Trans	**0.1171**	**7.20%**	**16.79%**	**0.1686**	**14.29%**	**28.65%**	**0.2084**	**18.78%**	**34.89%**	**0.2158**	**18.69%**	**32.73%**

Table 3: Main results on public amazon dataset

Method	Books		Music		Movie	
	NDCG	HR@20	NDCG	HR@20	NDCG	HR@20
Base	0.0270	4.71%	0.0631	13.39%	0.0433	10.45%
PPGN	0.0289	4.96%	0.0660	13.93%	0.0473	11.23%
MGNN	0.0311	5.12%	0.0672	14.14%	0.0465	11.03%
PCRec	0.0331	5.31%	0.0742	15.67%	0.0489	11.52%
BitGCF	0.0359	5.57%	0.0694	14.65%	0.0495	11.78%
BitGCF+	0.0381	5.78%	0.0719	15.29%	0.0509	12.02%
H^3Trans	**0.0399**	**5.97%**	**0.0761**	**16.01%**	**0.0524**	**12.33%**

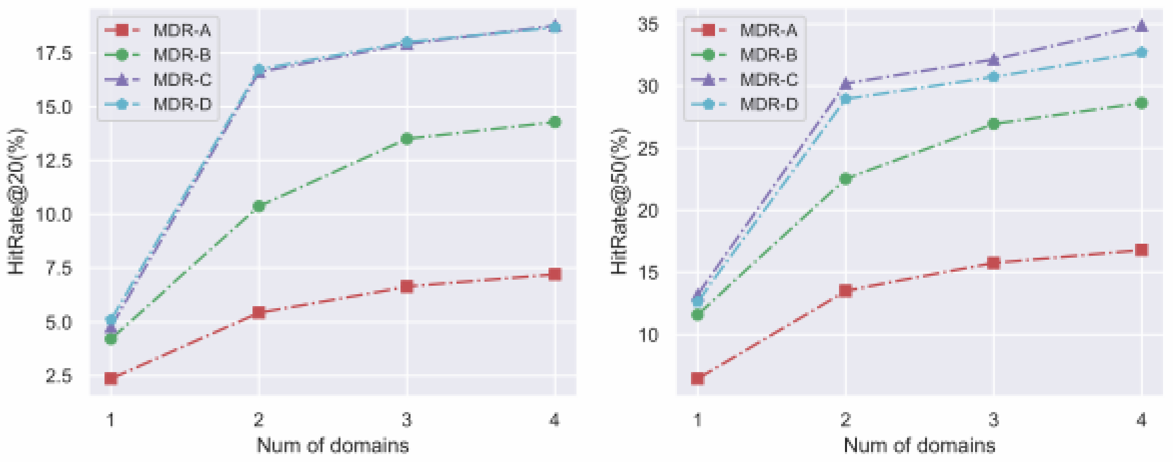


Figure 3: Performance comparison over different number of domains in MDR

Experiments

Table 4: **Ablation study on product dataset. Methods refer to different variants of** H^3Trans.

Method	MDR-A			MDR-B			MDR-C			MDR-D		
	Mrr	HR@20	HR@50	Mrr	HR@20	HR@50	Mrr	HR@20	HR@50	Mrr	HR@20	HR@50
Vanilla	0.0544	3.68%	8.11%	0.0699	5.34%	14.28%	0.1079	12.22%	21.34%	0.1428	10.67%	21.81%
HU	0.0750	5.08%	12.31%	0.1237	9.87%	20.71%	0.1712	16.15%	28.63%	0.1685	14.76%	25.85%
HU+	0.0894	5.56%	13.68%	0.1383	10.53%	23.08%	0.1848	17.01%	29.82%	0.1846	16.38%	28.48%
PHI	0.1016	6.35%	15.22%	0.1509	11.96%	24.52%	0.1887	17.58%	30.92%	0.1913	17.21%	29.80%
EHI	0.1051	6.53%	15.68%	0.1581	12.34%	25.54%	0.1958	17.93%	31.64%	0.1937	17.84%	30.62%
EHI+	**0.1171**	**7.20%**	**16.79%**	**0.1686**	**14.29%**	**28.65%**	**0.2084**	**18.78%**	**34.89%**	**0.2158**	**18.69%**	**32.73%**

Experiments

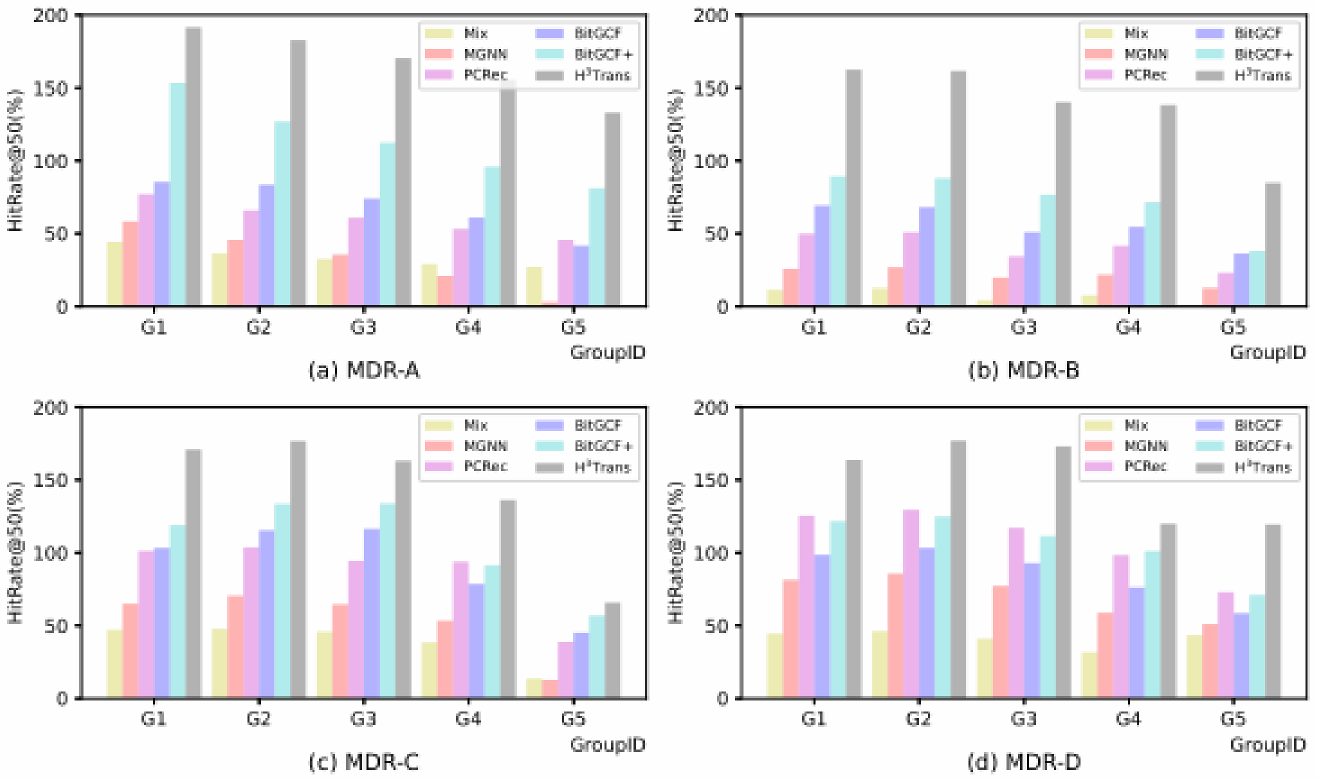


Figure 4: Performance comparison over different user groups (percentage increase relative to Base model)