Decoupled Side Information Fusion for Sequential Recommendation

Yueqi Xie*
HKUST
yxieay@connect.ust.hk

Peilin Zhou*
Upstage
zhoupalin@gmail.com

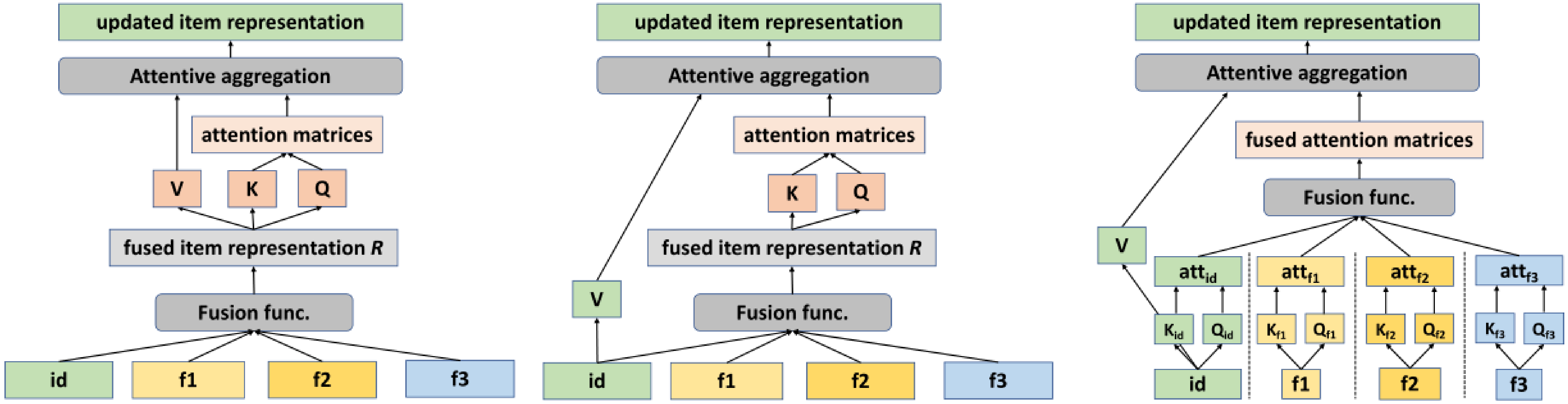
Sunghun Kim
HKUST
hunkim@ust.hk

code:https://github.com/AIM-SE/DIF-SR.

SIGIR2022

Reported by Minqin Li

Introduction



(a) SASRec$_F$. (b) NOVA-SR. (c) DIF-SR.

Figure 3: The comparison of item representation learning process of various solutions. (a) SASRec$_F$: SASRec$_F$ fuses side information into item representation and uses the fused item representation to calculate key, query and value. (b) NOVA-SR: NOVA-SR uses the fused item representation for the calculation of key and query, and keeps value non-invasive. (c) DIF-SR: Instead of early fusion to get fused item representation, the proposed DIF-SR decouples the attention calculation process of various side information to generate fused attention matrices for higher representation power, avoidance of mixed correlation, and flexible training gradient.

Method

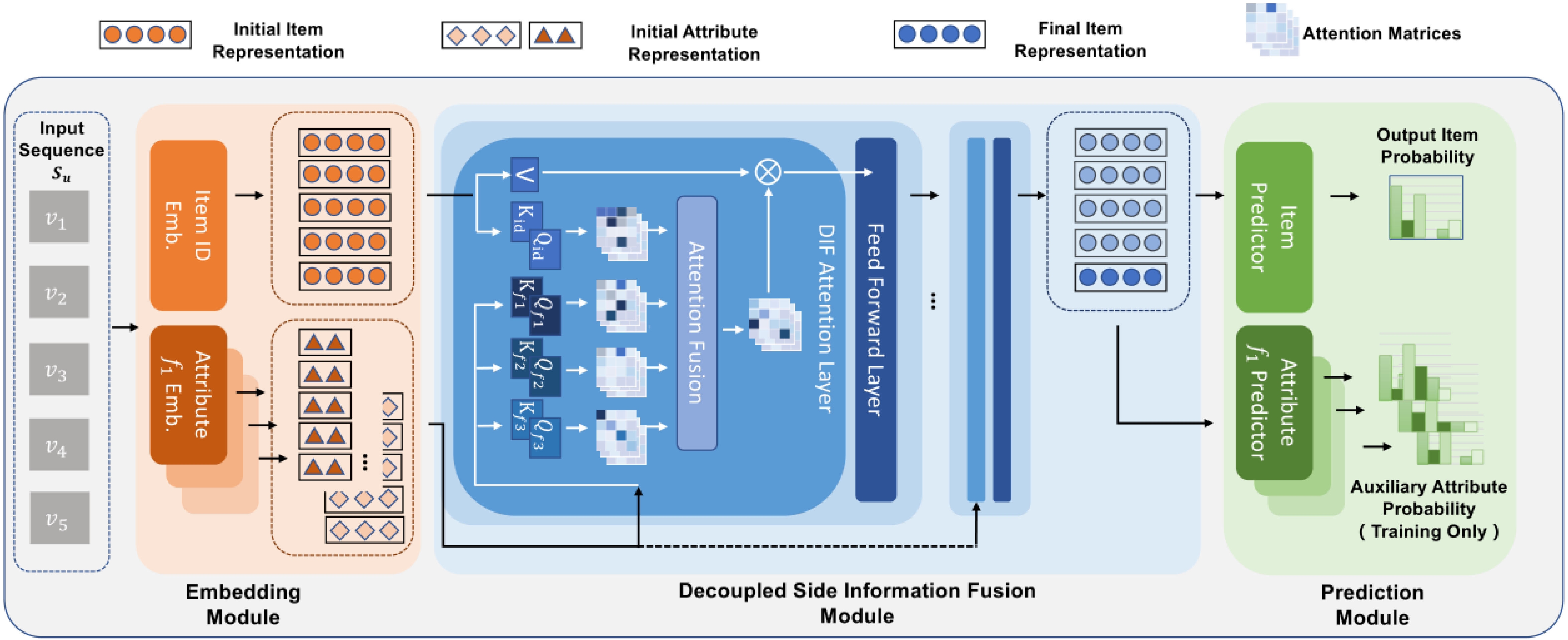


Figure 2: Overview of the proposed network.

Method

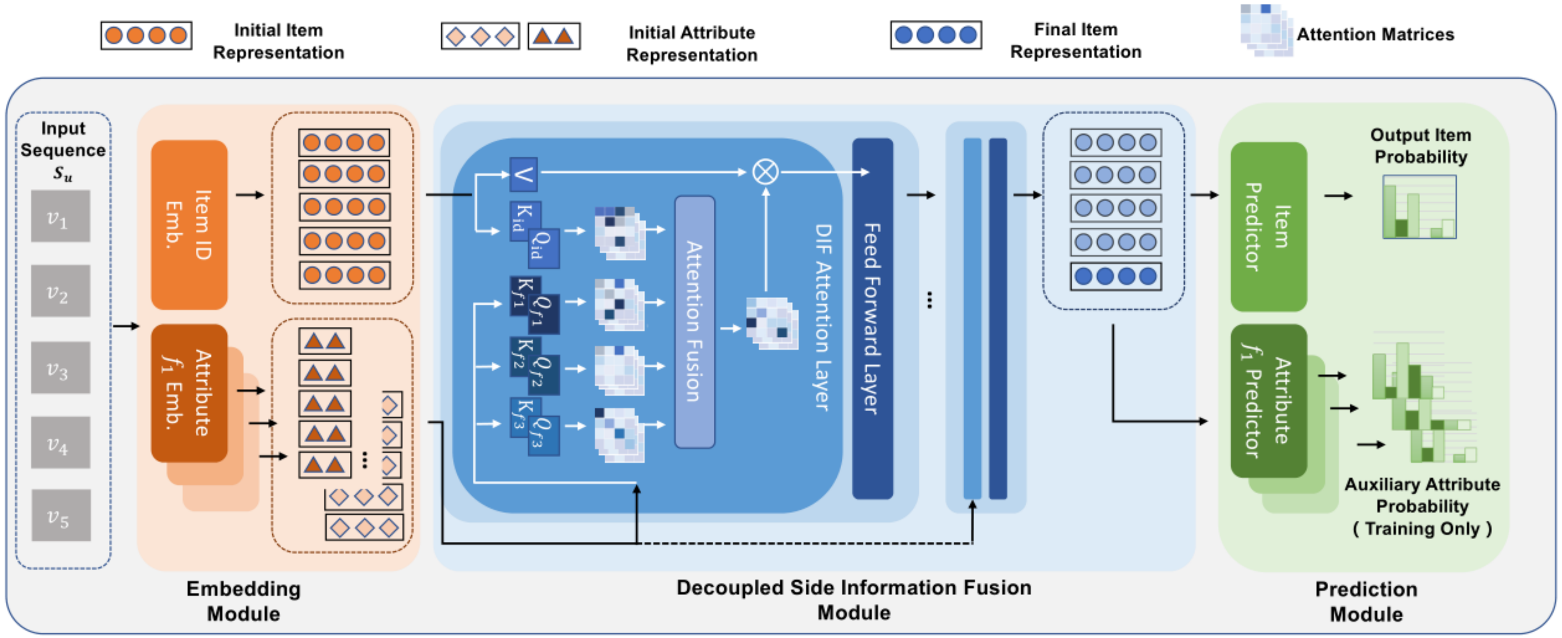


Figure 2: Overview of the proposed network.

Input: $\mathbb{S}_u = [v_1, v_2, \ldots, v_n]$

$$v_i = (I_i, f_i^{(1)}, \ldots, f_i^{(p)})$$

$$\begin{aligned} E^{ID} &= \mathcal{E}_{id}([I_1, I_2, \ldots, I_n]), \\ E^{f1} &= \mathcal{E}_{f1}([f_1^{(1)}, f_2^{(1)}, \ldots, f_n^{(1)}]), \\ &\cdots \\ E^{fp} &= \mathcal{E}_{fp}([f_1^{(p)}, f_2^{(p)}, \ldots, f_n^{(p)}]), \end{aligned} \tag{1}$$

$$R_{i+1}^{(ID)} = \mathrm{LN}(\mathrm{FFN}(\mathrm{DIF}(R_i^{(ID)}, E^{f1}, \ldots, E^{fp}))) \tag{2}$$

$$R_1^{(ID)} = E^{ID} \tag{3}$$

Method

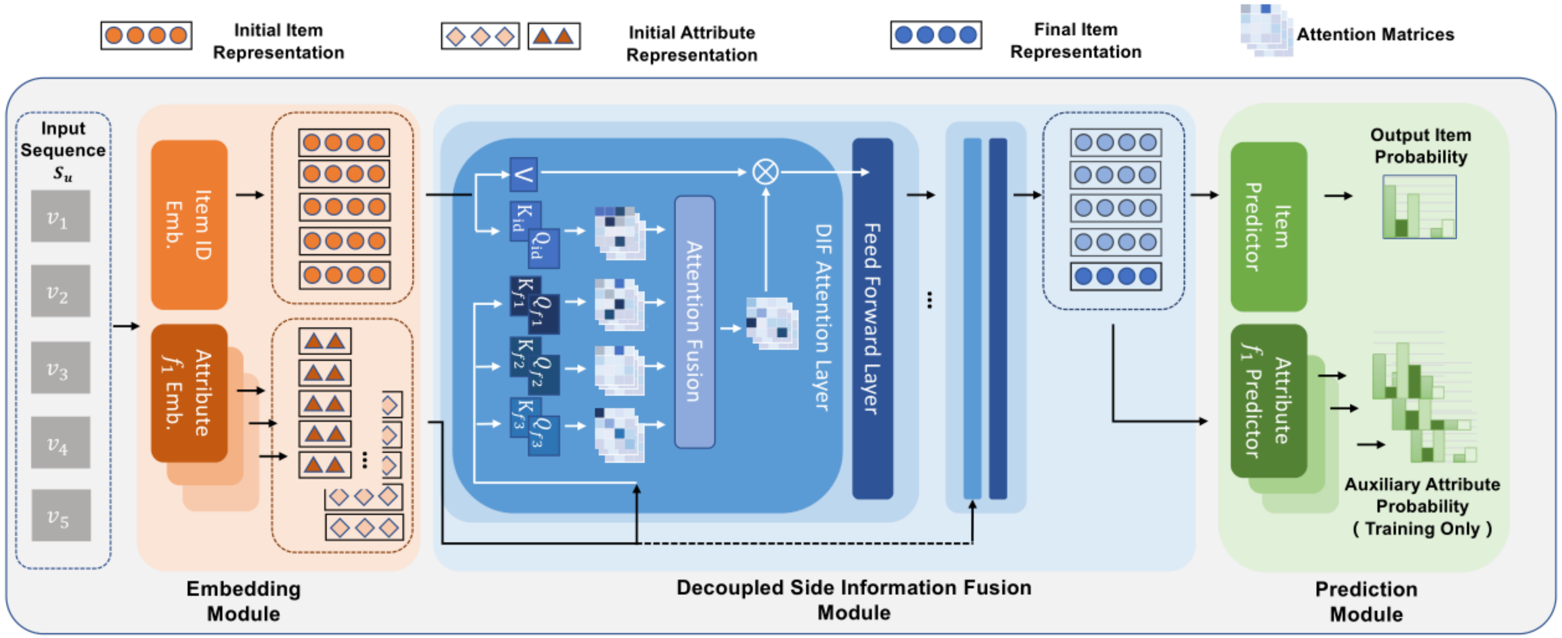


Figure 2: Overview of the proposed network.

$$\mathrm{att}_{ID}^{i} = (R^{(ID)} W_Q^{i})(R^{(ID)} W_K^{i})^{\top} \quad (4)$$

$$\begin{aligned}
\mathrm{att}_{f1}^{i} &= (E^{f1} W_Q^{(f1)i})(E^{f1} W_K^{(f1)i})^{\top}, \\
&\dots, \\
\mathrm{att}_{fp}^{i} &= (E^{fp} W_Q^{(fp)i})(E^{fp} W_K^{(fp)i})^{\top}.
\end{aligned} \quad (5)$$

Method

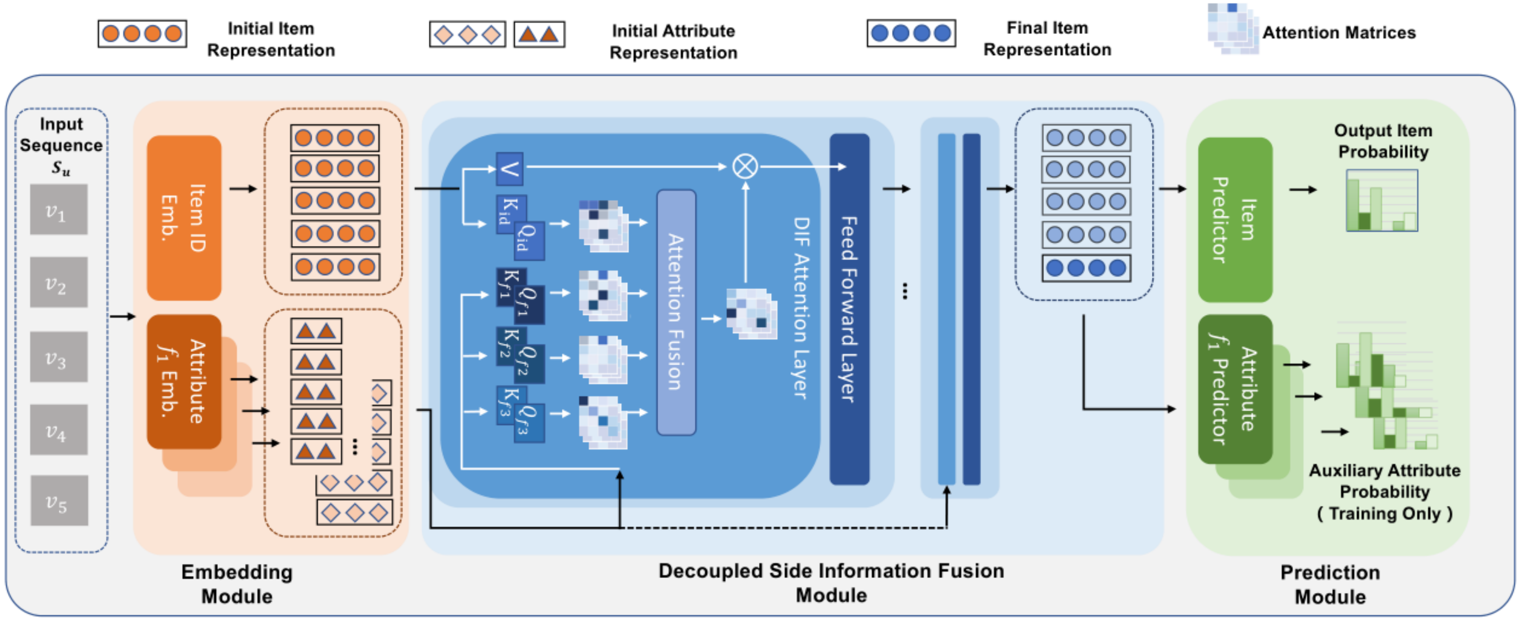


Figure 2: Overview of the proposed network.

① $$\mathcal{F}_{\text{add}}(f_1, \ldots, f_m) = \sum_{i=1}^{m} f_i$$

② $$\mathcal{F}_{\text{concat}}(f_1, \ldots, f_m) = \boldsymbol{FC}(f_1 \odot \cdots \odot f_m)$$

③ $$\mathcal{F}_{\text{gating}}(f_1, \ldots, f_m) = \sum_{i=1}^{m} G^{(i)} f_i$$

$$G = \sigma(FW^F)$$

$$\text{DIF\_att}^i = \mathcal{F}(\text{att}^i_{ID}, \text{att}^i_{f1}, \ldots, \text{att}^i_{fp}),$$
$$\text{DIF\_head}^i = \sigma\left(\frac{\text{DIF\_att}^i}{\sqrt{d}}\right)(R^{(ID)} W^i_V). \quad (6)$$

Method

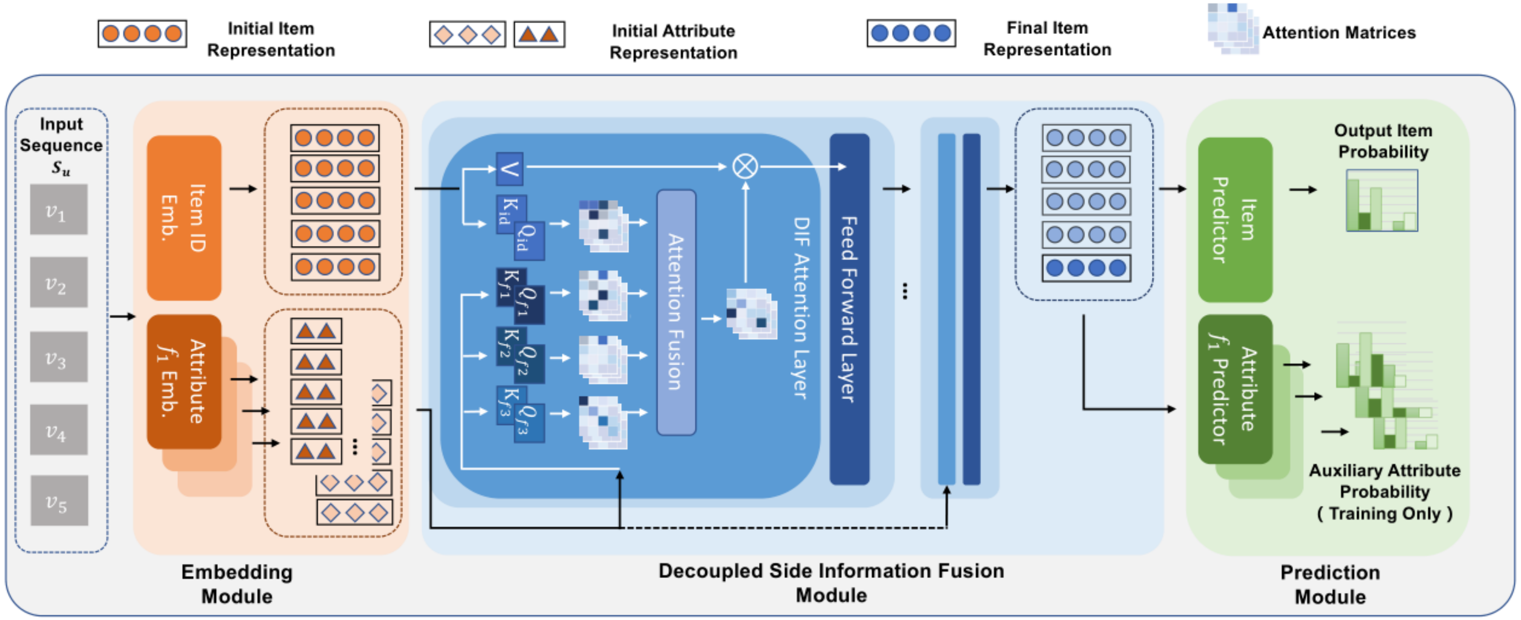


Figure 2: Overview of the proposed network.

$$\hat{y} = \mathrm{softmax}(M_{id} R_L^{(ID)} [n]^\top) \quad (7)$$

$$\hat{y}^{(fj)} = \sigma(W_{fj} R_L^{(ID)} [n]^\top + b_{fj}) \quad (8)$$

$$L_{id} = -\sum_{i=1}^{|\mathcal{I}|} y_i \log(\hat{y}_i) \quad (9)$$

$$L_{fj} = -\sum_{i=1}^{|fj|} y_i^{(fj)} \log(\hat{y}_i^{fj}) + (1 - y_i^{(fj)}) \log(1 - \hat{y}_i^{(fj)}) \quad (10)$$

$$L = L_{id} + \lambda \Sigma_{j=1}^{p} L_{fj} \quad (11)$$

Experiments

Table 1: Statistics of the datasets after preprocessing.

Dataset	Beauty	Sports	Toys	Yelp
# Users	22,363	35,598	19,412	30,499
# Items	12,101	18,357	11,924	20,068
# Avg. Actions / User	8.9	8.3	8.6	10.4
# Avg. Actions / Item	16.4	16.1	14.1	15.8
# Actions	198,502	296,337	167,597	317,182
Sparsity	99.93%	99.95%	99.93%	99.95%

Experiments

Table 2: Overall performance. Bold scores represent the highest results of all methods. Underlined scores stand for the highest results from previous methods.

Dataset	Metric	GRU4Rec	Caser	BERT4Rec	$GRU4Rec_F$	SASRec	$SASRec_F$	S^3Rec	NOVA	ICAI	DIF-SR
Beauty	Recall@10	0.0530	0.0474	0.0529	0.0587	0.0828	0.0719	0.0868	0.0887	0.0879	**0.0908**
	Recall@20	0.0839	0.0731	0.0815	0.0902	0.1197	0.1013	0.1236	0.1237	0.1231	**0.1284**
	NDCG@10	0.0266	0.0239	0.0237	0.0290	0.0371	0.0414	0.0439	0.0439	0.0439	**0.0446**
	NDCG@20	0.0344	0.0304	0.0309	0.0369	0.0464	0.0488	0.0531	0.0527	0.0528	**0.0541**
Sports	Recall@10	0.0312	0.0227	0.0295	0.0394	0.0526	0.0435	0.0517	0.0534	0.0527	**0.0556**
	Recall@20	0.0482	0.0364	0.0465	0.0610	0.0773	0.0640	0.0758	0.0759	0.0762	**0.0800**
	NDCG@10	0.0157	0.0118	0.0130	0.0199	0.0233	0.0235	0.0249	0.0250	0.0243	**0.0264**
	NDCG@20	0.0200	0.0153	0.0173	0.0253	0.0295	0.0286	0.0310	0.0307	0.0302	**0.0325**
Toys	Recall@10	0.0370	0.0361	0.0533	0.0492	0.0831	0.0733	0.0967	0.0978	0.0972	**0.1013**
	Recall@20	0.0588	0.0566	0.0787	0.0767	0.1168	0.1052	0.1349	0.1322	0.1303	**0.1382**
	NDCG@10	0.0184	0.0186	0.0234	0.0246	0.0375	0.0417	0.0475	0.0480	0.0478	**0.0504**
	NDCG@20	0.0239	0.0238	0.0297	0.0316	0.0460	0.0497	0.0571	0.0567	0.0561	**0.0597**
Yelp	Recall@10	0.0361	0.0380	0.0524	0.0361	0.0650	0.0413	0.0589	0.0681	0.0663	**0.0698**
	Recall@20	0.0592	0.0608	0.0756	0.0578	0.0928	0.0675	0.0902	0.0964	0.0940	**0.1003**
	NDCG@10	0.0184	0.0197	0.0327	0.0182	0.0401	0.0216	0.0338	0.0412	0.0400	**0.0419**
	NDCG@20	0.0243	0.0255	0.0385	0.0236	0.0471	0.0282	0.0416	0.0483	0.0470	**0.0496**

Experiments

Table 3: Performance comparison of self-attention based sequential models with their DIF & AAP incorporated version on Beauty, Sports and Toys datasets.

Dataset	Metric	DIF-SASRec w/o	DIF-SASRec w	DIF-BERT4Rec w/o	DIF-BERT4Rec w	Improv. DIF-SASRec	Improv. DIF-BERT4Rec
Beauty	Recall@10	0.0828	0.0908	0.0529	0.0579	9.66%	9.45%
	Recall@20	0.1197	0.1284	0.0815	0.0915	7.27%	12.27%
	NDCG@10	0.0371	0.0446	0.0237	0.0279	20.22%	17.72%
	NDCG@20	0.0464	0.0541	0.0309	0.0363	16.59%	17.48%
Sports	Recall@10	0.0526	0.0556	0.0295	0.0394	5.70%	33.56%
	Recall@20	0.0773	0.0800	0.0465	0.0611	3.49%	31.40%
	NDCG@10	0.0233	0.0264	0.0130	0.0198	13.30%	52.31%
	NDCG@20	0.0295	0.0325	0.0173	0.0252	10.17%	45.66%
Toys	Recall@10	0.0831	0.1013	0.0533	0.0599	21.90%	12.38%
	Recall@20	0.1168	0.1382	0.0787	0.0851	18.32%	8.13%
	NDCG@10	0.0375	0.0504	0.0234	0.0324	34.40%	38.46%
	NDCG@20	0.0460	0.0597	0.0297	0.0387	29.78%	30.30%

Experiments

Table 4: Ablation study of DIF and AAP on Yelp, Sports and Beauty datasets.

Settings		Yelp		Sports		Beauty	
DIF	AAP	Recall@20	+Δ	Recall@20	+Δ	Recall@20	+Δ
✗	✗	0.0663	-	0.0621	-	0.0996	-
✗	✓	0.0663	+0%	0.0754	+21.42%	0.0991	-0.50%
✓	✗	0.0968	+46.00%	0.0767	+23.51%	0.1240	+24.50%
✓	✓	0.1003	+51.28%	0.0800	+28.82%	0.1284	+28.92%

Table 5: Performance comparison of using different kinds of side information on Yelp dataset.

Side-info	Recall@10	Recall@20	NDCG@10	NDCG@20
Position	0.0655	0.0954	0.0405	0.048
Position + Categories	0.0698	0.1003	0.0419	0.0496
Position + City	0.0691	0.1001	0.0415	0.0493
Position + Categories + City	**0.0699**	**0.1010**	**0.0421**	**0.0499**

Experiments

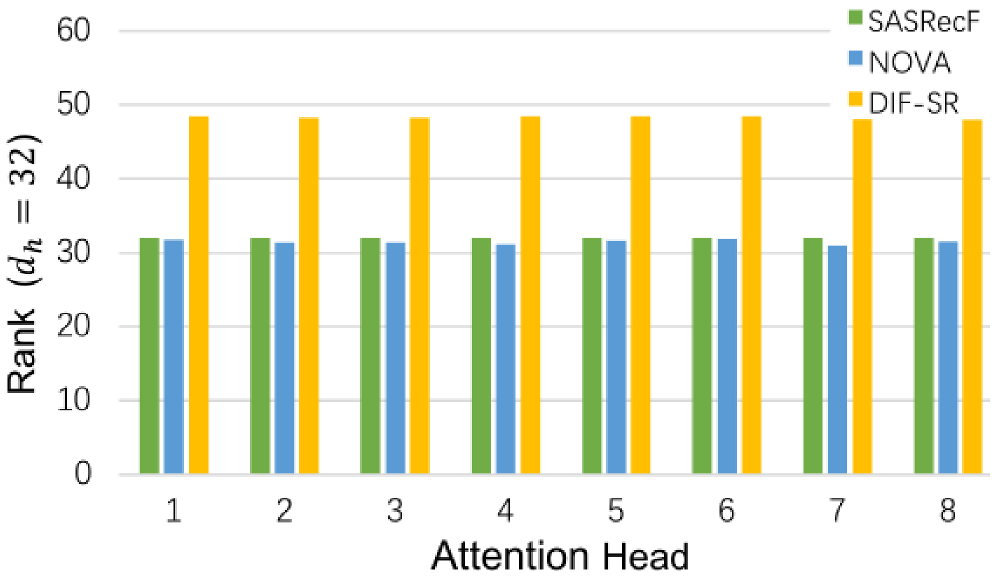


Figure 1: Rank of attention matrices: A comparison of the average rank of attention score matrices of early-integrated embedding based solutions, i.e., SASRecF and NOVA, and our proposed DIF-SR. The early-integration of embeddings leads to lower rank of the attention matrices and limits the expressiveness.

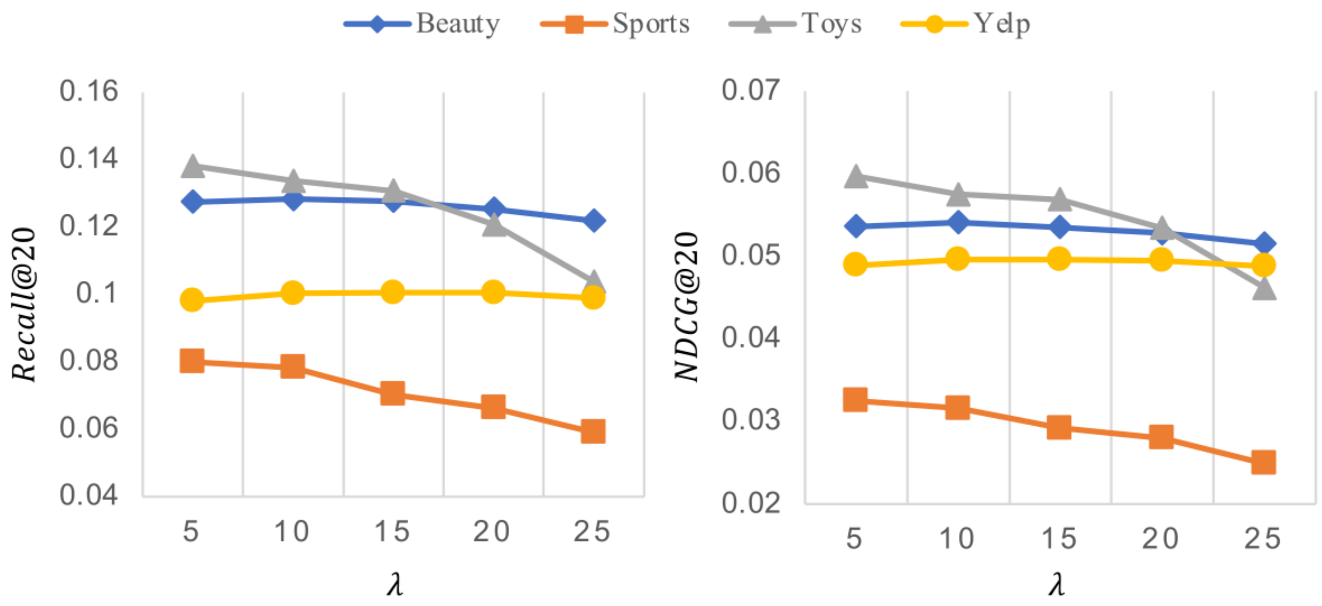


Figure 4: Effect of balance parameter λ.

Experiments

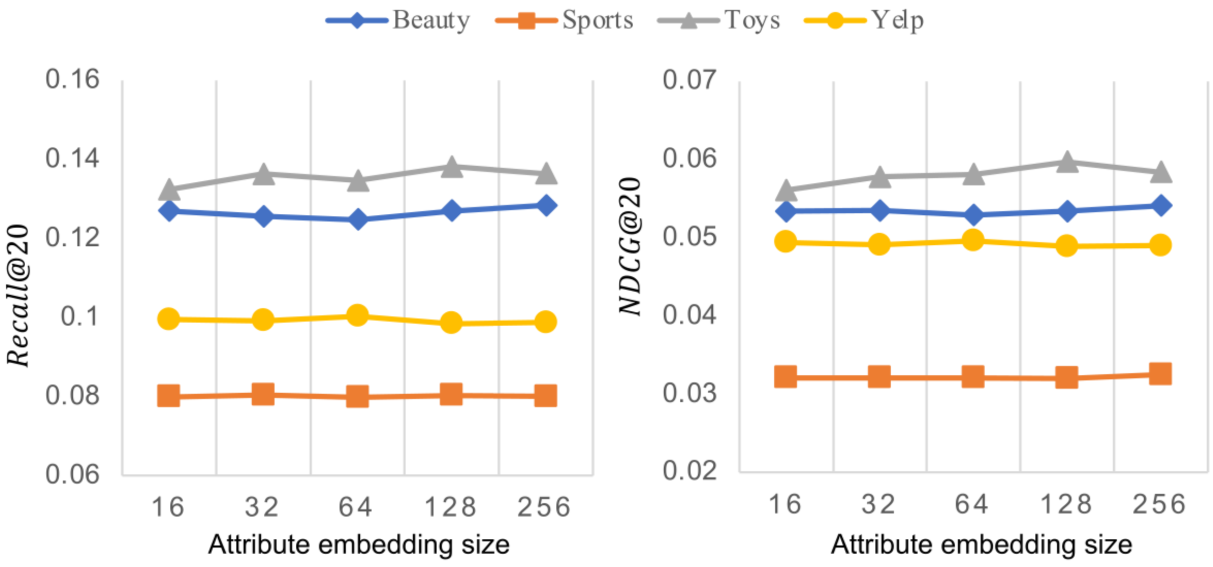


Figure 5: Effect of attribute embedding size d_f..

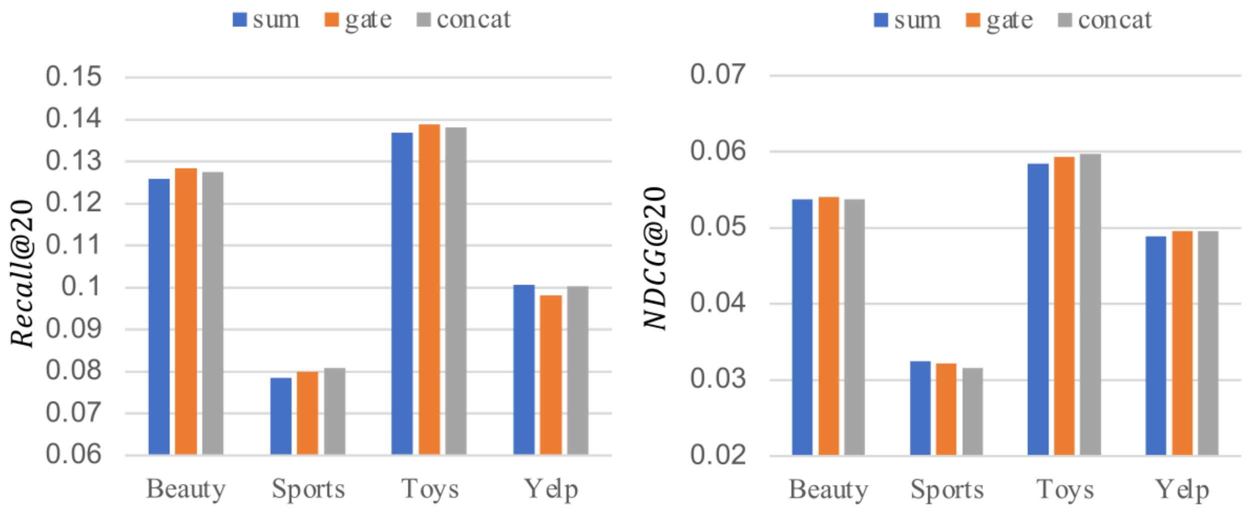


Figure 6: Effect of fusion functions $\mathcal{F}$.

Experiments

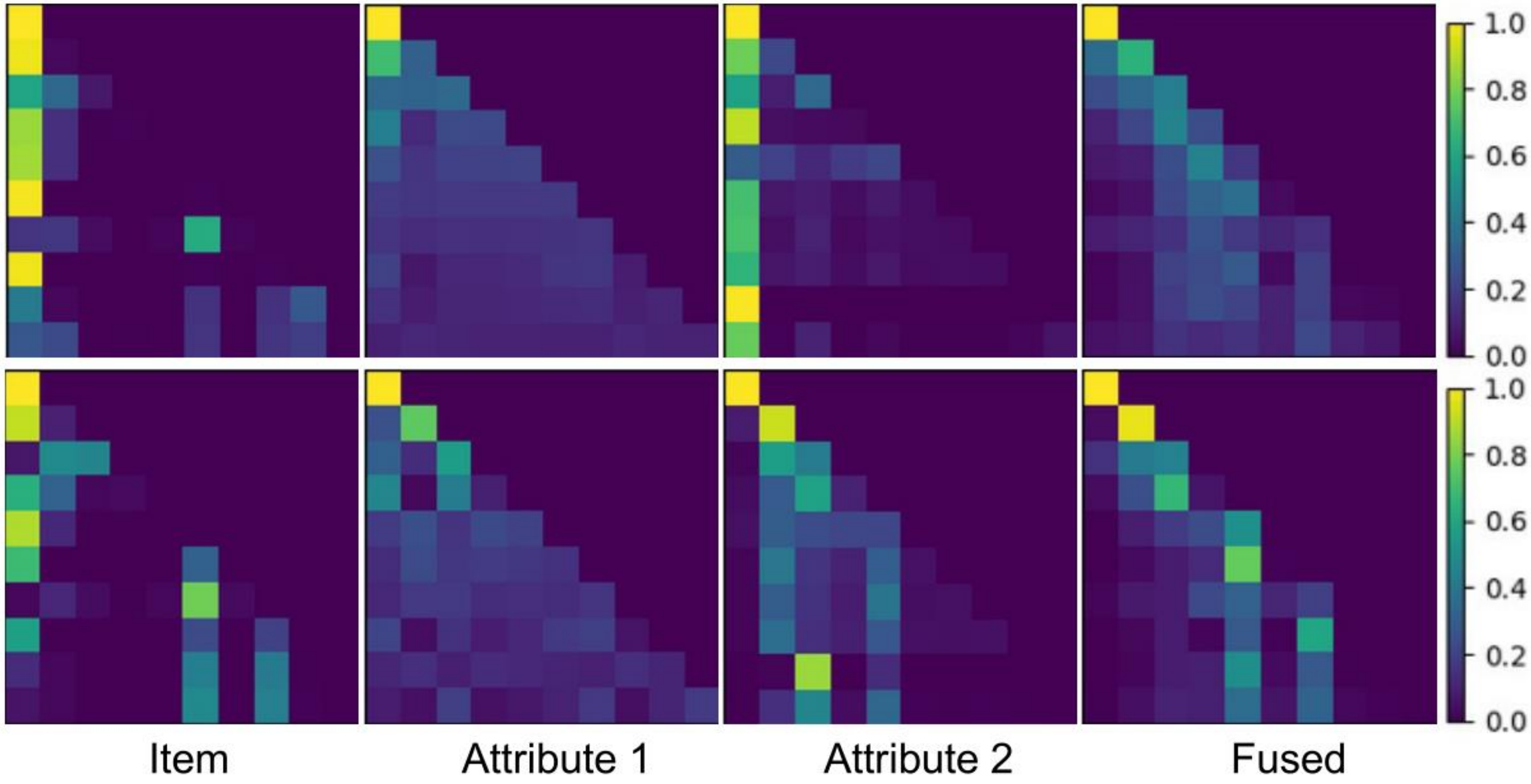


Figure 7: Visualization of sampled attention matrices.

Thanks

```
15 Nov 16:18    INFO  best valid : {'recall@3': 0.0518, 'recall@5': 0.0768, 'recall@10': 0.113, 'recall@20': 0.1585, 'ndcg@3': 0.0333, 'ndcg@5': 0.0435, 'ndcg@10': 0.0553, 'ndcg@20': 0.0667}
15 Nov 16:18    INFO  test result: {'recall@3': 0.0402, 'recall@5': 0.0579, 'recall@10': 0.088, 'recall@20': 0.1268, 'ndcg@3': 0.0266, 'ndcg@5': 0.0338, 'ndcg@10': 0.0435, 'ndcg@20': 0.0533}
```

```
15 Nov 21:08    INFO  best valid : {'recall@3': 0.0295, 'recall@5': 0.0432, 'recall@10': 0.0672, 'recall@20': 0.098, 'ndcg@3': 0.0181, 'ndcg@5': 0.0238, 'ndcg@10': 0.0315, 'ndcg@20': 0.0393}
15 Nov 21:08    INFO  test result: {'recall@3': 0.024, 'recall@5': 0.0343, 'recall@10': 0.053, 'recall@20': 0.0762, 'ndcg@3': 0.0148, 'ndcg@5': 0.019, 'ndcg@10': 0.025, 'ndcg@20': 0.0308}
7070.024189710617
(lmq) bigdata2@bigdata2:~/lmq/DIF-SR-main$
```

```
16 Nov 10:44    INFO  best valid : {'recall@3': 0.0559, 'recall@5': 0.0821, 'recall@10': 0.1191, 'recall@20': 0.1632, 'ndcg@3': 0.0362, 'ndcg@5': 0.047, 'ndcg@10': 0.0589, 'ndcg@20': 0.07}
16 Nov 10:44    INFO  test result: {'recall@3': 0.0487, 'recall@5': 0.0688, 'recall@10': 0.0992, 'recall@20': 0.135, 'ndcg@3': 0.0318, 'ndcg@5': 0.0401, 'ndcg@10': 0.0499, 'ndcg@20': 0.0589}
4579.508291959763
(lmq) bigdata2@bigdata2:~/lmq/DIF-SR-main$
```

```
17 Nov 12:11    INFO  best valid : {'recall@3': 0.0276, 'recall@5': 0.0411, 'recall@10': 0.0639, 'recall@20': 0.0969, 'ndcg@3': 0.0167, 'ndcg@5': 0.0223, 'ndcg@10': 0.0296, 'ndcg@20': 0.0379}
17 Nov 12:11    INFO  test result: {'recall@3': 0.027, 'recall@5': 0.0377, 'recall@10': 0.0579, 'recall@20': 0.0824, 'ndcg@3': 0.0167, 'ndcg@5': 0.0212, 'ndcg@10': 0.0277, 'ndcg@20': 0.0338}
3215.3000197410583
(lmq) bigdata2@bigdata2:~/lmq/DIF-SR-main$
```