Aligned Side Information Fusion Method for Sequential Recommendation

Shuhan Wang
hanshu.wsh@antgroup.com
Ant Group
Hangzhou, China

Bin Shen
ringo.sb@antgroup.com
Ant Group
Hangzhou, China

Xu Min
minxu.mx@antgroup.com
Ant Group
Hangzhou, China

Yong He
heyong.h@antgroup.com
Ant Group
Hangzhou, China

Xiaolu Zhang
yueyin.zxl@antgroup.com
Ant Group
Hangzhou, China

Liang Zhang
zhuyue.zl@antgroup.com
Ant Group
Hangzhou, China

Jun Zhou
jun.zhoujun@antgroup.com
Ant Group
Hangzhou, China

Linjian Mo
linyi01@antgroup.com
Ant Group
Hangzhou, China

code:none

WWW 2024

Reported by Minqin Li

Introduction

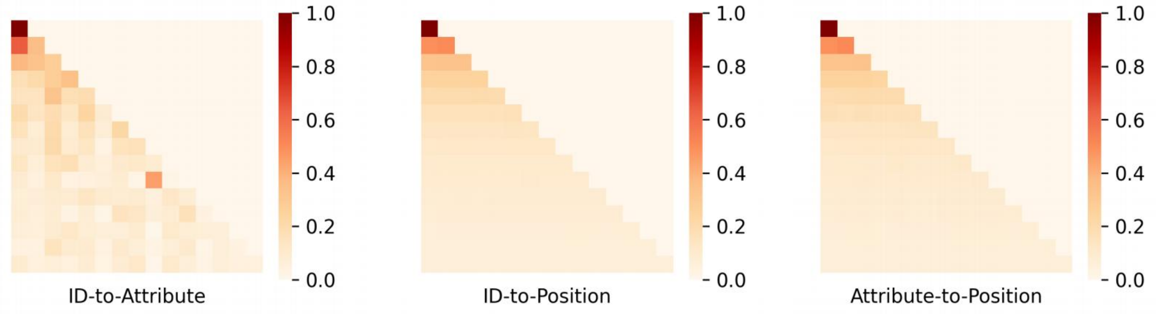
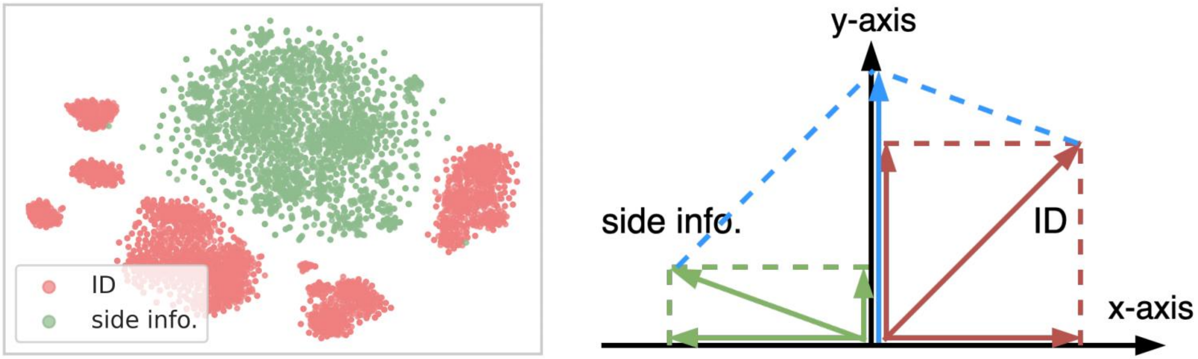


Figure 1: Visualization of attention scores in SASRec_F on Yelp dataset.



(a) Macroscopic perspective (b) Microscopic perspective

Figure 2: Visualization of information invasion.

Motivation:

Combining Side Information beyond IDs has become an important way to improve the performance in recommender systems.

Challenges:

1.Difficult to eliminate interference and learn meaningful signals from noisy correlations.

2.Difficult to avoid information invasion.

Introduction

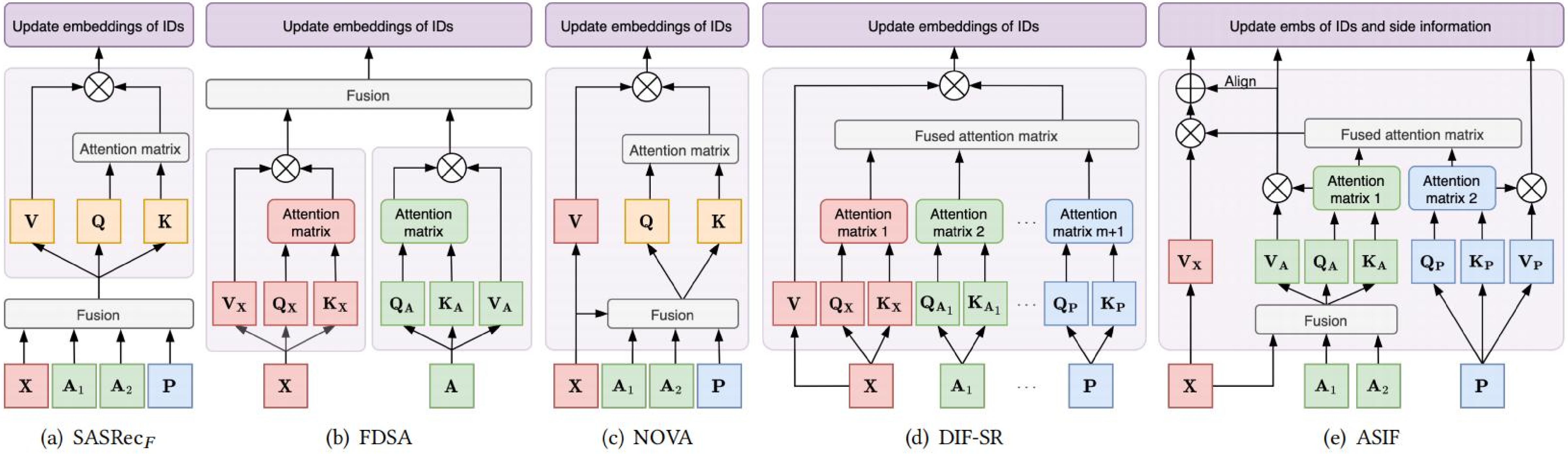


Figure 3: Single layer structure comparison of existing self-attention-based side information fusion approaches: SASRec$_F$ is early fusion, FDSA is late fusion, while NOVA, DIF-SR and ASIF is hybrid fusion.

Method

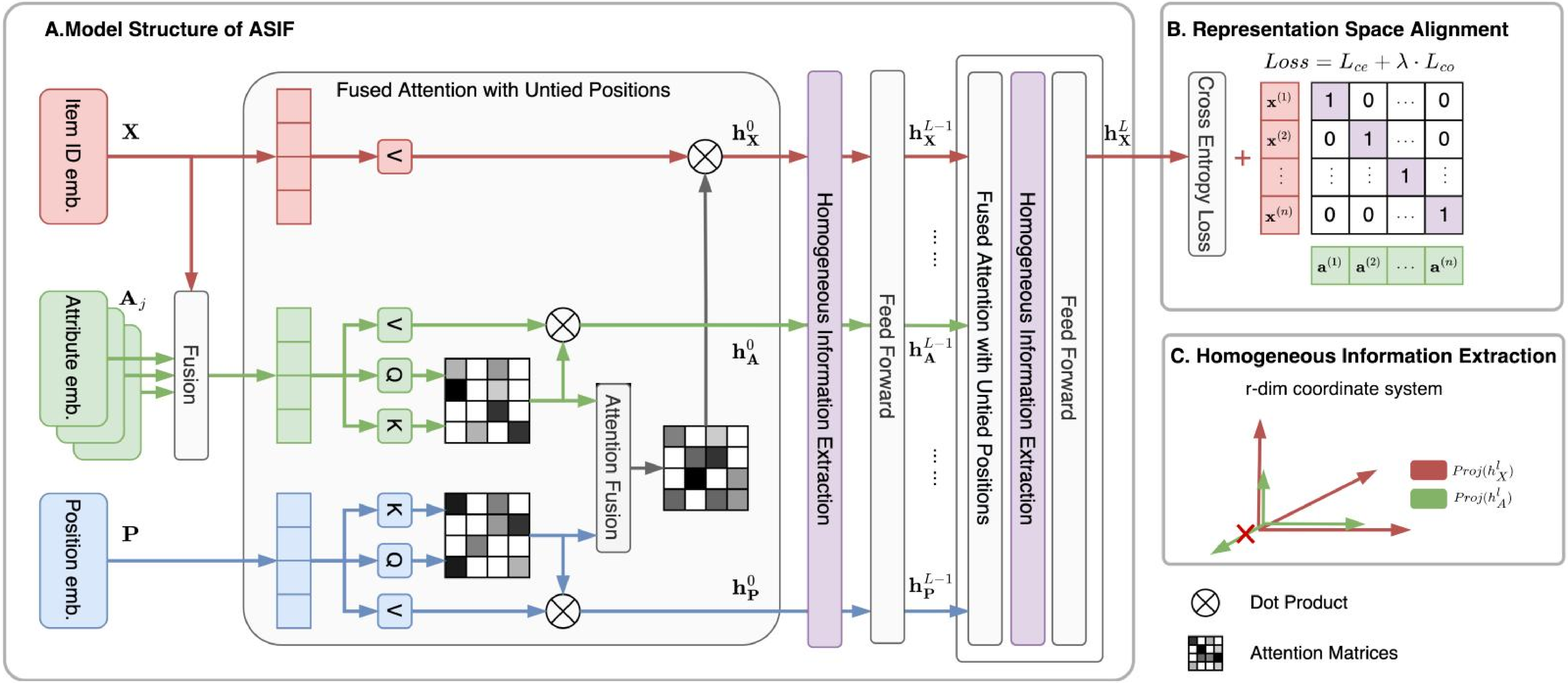
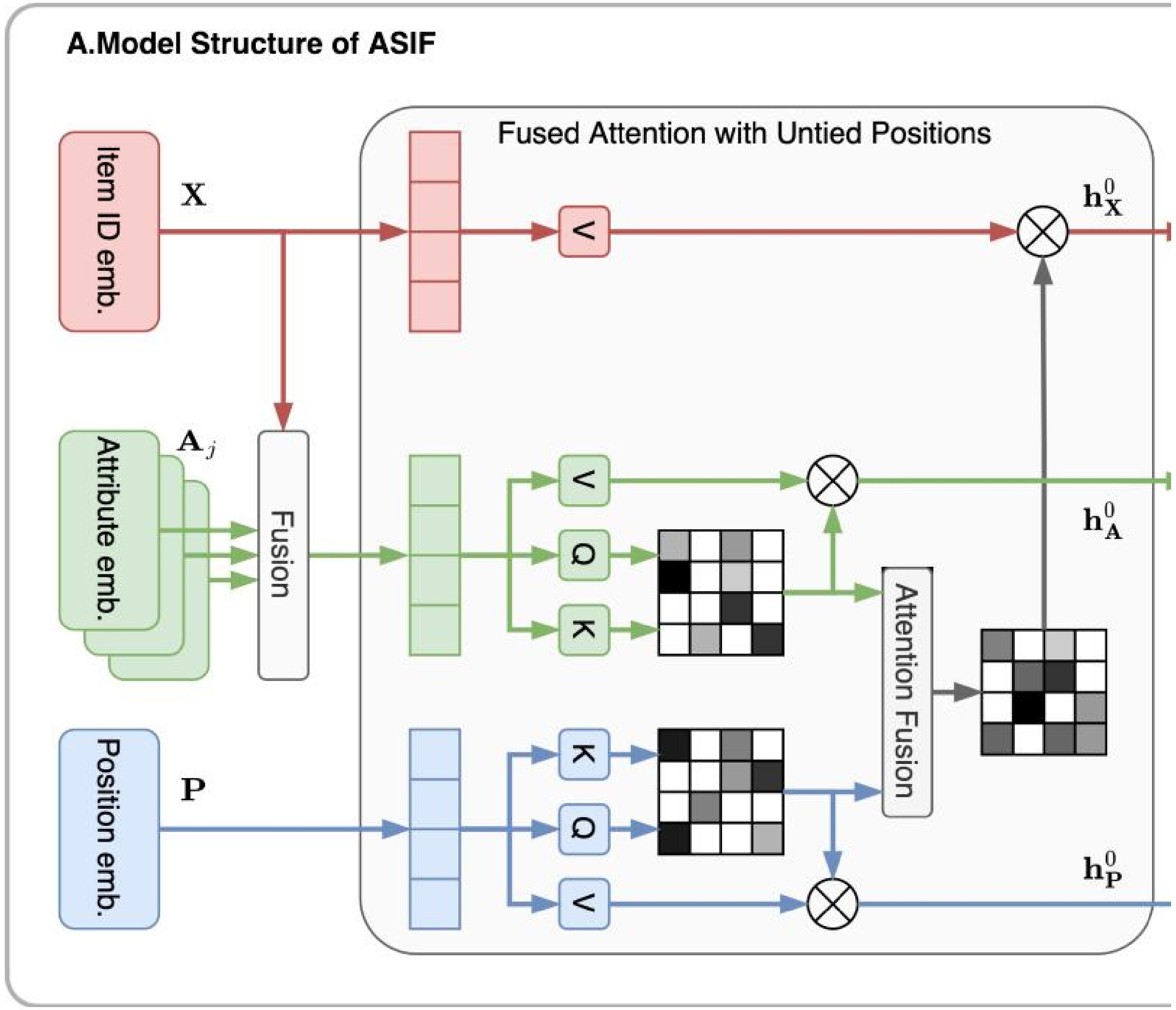


Figure 4: An overview of ASIF.

Method



$$\mathbf{C}_{\mathbf{XA}} = \mathcal{F}(\mathbf{X}, \mathbf{A})\mathbf{W}_{q,1}\mathbf{W}_{k,1}^{T}\mathcal{F}(\mathbf{X}, \mathbf{A})^{T} \quad (1)$$

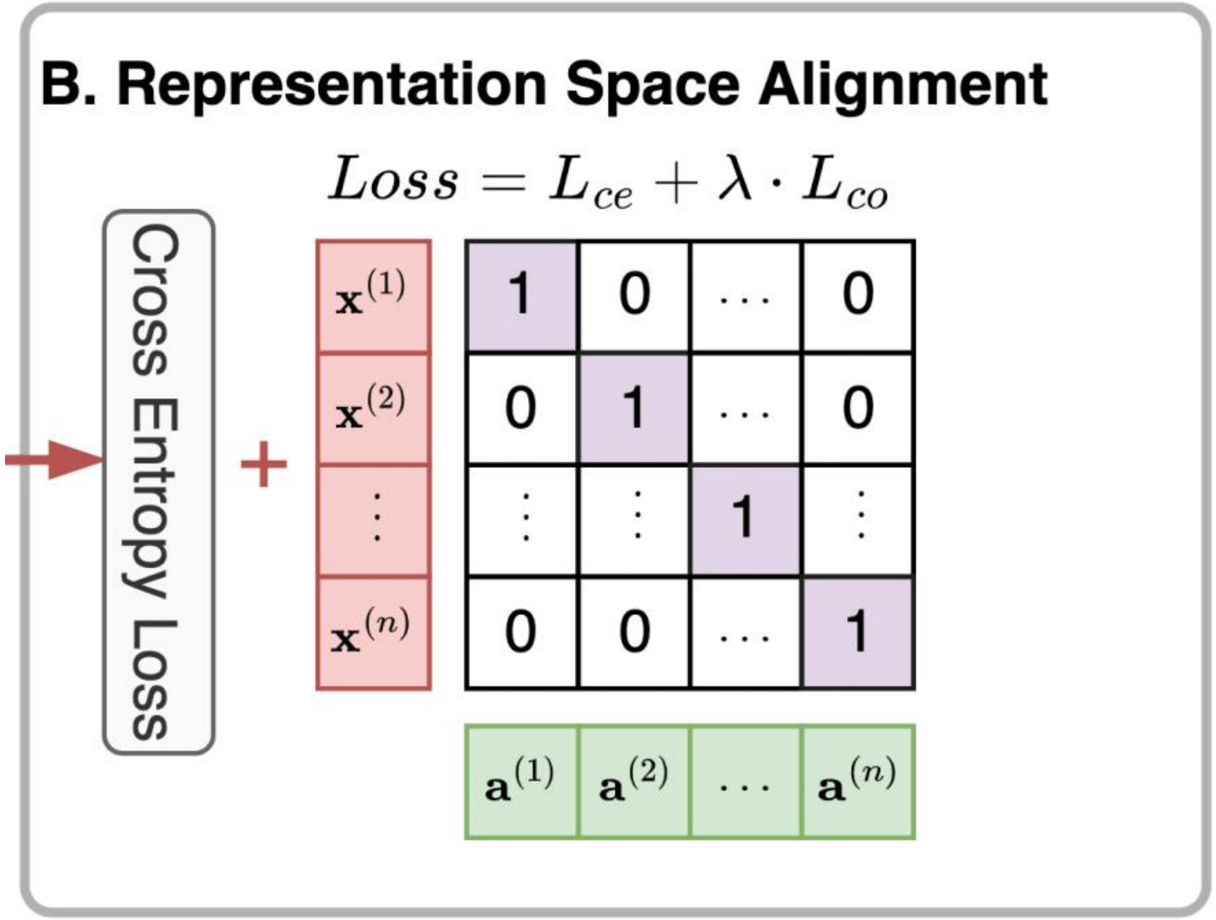
$$\mathbf{C}_{\mathbf{P}} = \mathbf{P}\mathbf{W}_{q,2}\mathbf{W}_{k,2}^{T}\mathbf{P}^{T} \quad (2)$$

$$\begin{aligned}\mathbf{h}_{\mathbf{X}} &= \mathrm{FusedAttention}(\mathbf{X}, \mathbf{A}_1, \cdots, \mathbf{A}_m, \mathbf{P}) \\ &= \mathrm{Softmax}\left(\frac{\mathbf{C}_{\mathbf{XA}} + \mathbf{C}_{\mathbf{P}}}{\sqrt{d_h}}\right)\mathbf{X}\mathbf{W}_{v,1},\end{aligned} \quad (3)$$

$$\begin{aligned}\mathbf{h}_{\mathbf{A}} &= \mathrm{FusedAttention}(\mathbf{X}, \mathbf{A}_1, \cdots, \mathbf{A}_m) \\ &= \mathrm{Softmax}\left(\frac{\mathbf{C}_{\mathbf{XA}}}{\sqrt{d_h}}\right)\mathcal{F}(\mathbf{X}, \mathbf{A})\mathbf{W}_{v,2},\end{aligned} \quad (4)$$

$$\mathbf{h}_{\mathbf{P}} = \mathrm{FusedAttention}(\mathbf{P}) = \mathrm{Softmax}\left(\frac{\mathbf{C}_{\mathbf{P}}}{\sqrt{d_h}}\right)\mathbf{P}\mathbf{W}_{v,3} \quad (5)$$

Method



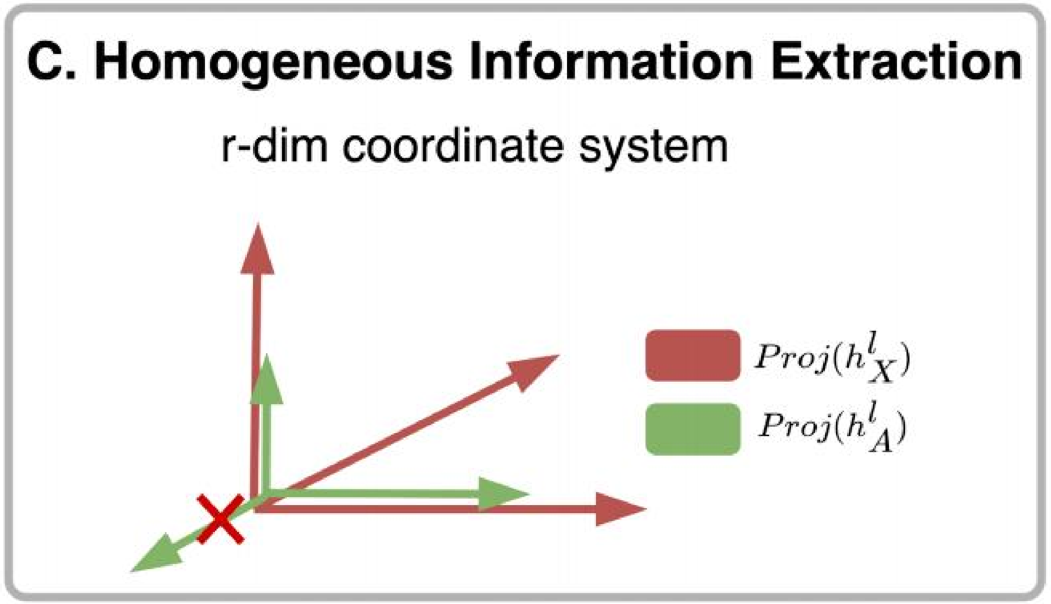
$$X = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \vdots \\ \mathbf{x}^{(n)} \end{bmatrix}, \quad A = \sum_{j=1}^{m} A_j = \begin{bmatrix} \mathbf{a}^{(1)} \\ \mathbf{a}^{(2)} \\ \vdots \\ \mathbf{a}^{(n)} \end{bmatrix} \tag{6}$$

$$\widetilde{X} = \begin{bmatrix} \mathbf{x}^{(1)}/\|\mathbf{x}^{(1)}\| \\ \mathbf{x}^{(2)}/\|\mathbf{x}^{(2)}\| \\ \vdots \\ \mathbf{x}^{(n)}/\|\mathbf{x}^{(n)}\| \end{bmatrix}, \quad \widetilde{A} = \begin{bmatrix} \mathbf{a}^{(1)}/\|\mathbf{a}^{(1)}\| \\ \mathbf{a}^{(2)}/\|\mathbf{a}^{(2)}\| \\ \vdots \\ \mathbf{a}^{(n)}/\|\mathbf{a}^{(n)}\| \end{bmatrix} \tag{7}$$

$$\widehat{Y}_X = \text{Softmax}\left(\widetilde{X}\widetilde{A}^T/\tau\right), \quad \widehat{Y}_A = \text{Softmax}\left(\widetilde{A}\widetilde{X}^T/\tau\right) \tag{8}$$

$$L_{co} = -\frac{1}{2N}\sum_{i=1}^{N}\sum\left(Y^i \odot \log \widehat{Y}_X^i + Y^i \odot \log \widehat{Y}_A^i\right) \tag{9}$$

Method



$$\mathrm{Proj}(\mathbf{h}_\mathrm{X}) = \mathbf{h}_\mathrm{X}\mathbf{Q}, \quad \mathrm{Proj}(\mathbf{h}_\mathrm{A}) = \mathbf{h}_\mathrm{A}\mathbf{Q}, \tag{10}$$

$$\widetilde{\mathrm{Proj}}(\mathbf{h}_\mathrm{A}) = \phi\left(\mathrm{Proj}(\mathbf{h}_\mathrm{X}) \odot \mathrm{Proj}(\mathbf{h}_\mathrm{A})\right) \odot \mathrm{Proj}(\mathbf{h}_\mathrm{A}) \tag{11}$$

$$\mathbf{h}^*_\mathrm{A} = \widetilde{\mathrm{Proj}}(\mathbf{h}_\mathrm{A})\mathbf{Q}^T \tag{12}$$

$$\begin{aligned}\mathbf{h}_\mathrm{X} &= \mathrm{FusedAttention}(\mathbf{X}, \mathbf{A}_1, \cdots, \mathbf{A}_m, \mathbf{P}) + \mathbf{h}_\mathrm{A}{}^* \\ &= \mathrm{Softmax}\left(\frac{\mathbf{C}_\mathrm{XA} + \mathbf{C}_\mathrm{P}}{\sqrt{d_h}}\right)\mathbf{X}\mathbf{W}_{v,1} + \mathbf{h}_\mathrm{A}{}^*.\end{aligned} \tag{13}$$

$$\widehat{y} = \mathrm{Softmax}(\mathbf{h}^L_\mathrm{X} \cdot \mathbf{V}) \tag{14}$$

$$L_{ce} = -\frac{1}{N}\sum_{i=1}^{N} y^i \log \widehat{y}^i \tag{15}$$

$$\begin{aligned}L &= L_{ce} + \lambda \cdot L_{co} \\ &= -\frac{1}{N}\sum_{i=1}^{N}\left(y^i \log \widehat{y}^i + \frac{\lambda}{2}\sum\left(\mathrm{Y}^i \odot \log \widehat{\mathrm{Y}}^i_\mathrm{X} + \mathrm{Y}^i \odot \log \widehat{\mathrm{Y}}^i_\mathrm{A}\right)\right)\end{aligned} \tag{16}$$

Experiments

Table 1: Statistics of datasets.

Dataset	# Users	# Items	# Actions	# Avg. len
Yelp	30450	20039	316541	10.4
AliEC	34148	18654	290490	8.5
Beauty	22364	12102	198502	8.9
Industrial	33061	19873	290000	8.8

Experiments

Model	Yelp				AliEC				Beauty			
	H@10	H@20	N@10	N@20	H@10	H@20	N@10	N@20	H@10	H@20	N@10	N@20
Bert4Rec	0.0354	0.0580	0.0189	0.0246	0.0503	0.0756	0.0263	0.0327	0.0542	0.0793	0.0315	0.0378
Caser	0.0357	0.0573	0.0177	0.0231	0.0336	0.0522	0.0171	0.0218	0.0416	0.0672	0.0211	0.0275
GRU4Rec	0.0350	0.0579	0.0175	0.0232	0.0361	0.0567	0.0182	0.0234	0.0510	0.0766	0.0268	0.0333
SASRec	0.0647	0.0936	0.0398	0.0471	0.0903	0.1300	0.0449	0.0549	0.0861	0.1225	0.0424	0.0516
LightSANs	0.0658	0.0970	0.0402	0.0480	0.0942	0.1354	0.0470	0.0574	0.0871	0.1242	0.0441	0.0535
FMLP	0.0657	0.0935	0.0400	0.0470	0.0936	0.1346	0.0463	0.0566	0.0855	0.1190	0.0450	0.0534
GRU4Rec$_F$	0.0362	0.0605	0.0182	0.0243	0.0471	0.0743	0.0237	0.0305	0.0532	0.0820	0.0274	0.0347
SASRec$_F$	0.0467	0.0749	0.0249	0.0319	0.0719	0.1081	0.0383	0.0474	0.0776	0.1082	0.0447	0.0540
LightSANs$_F$	0.0641	0.0925	0.0390	0.0461	0.0944	0.1382	0.0469	0.0579	0.0880	0.1244	0.0448	0.0540
FMLP$_F$	0.0629	0.0884	0.0385	0.0448	0.0997	0.1431	0.0495	0.0604	0.0871	0.1220	0.0452	0.0540
CL4SRec	0.0666	0.0965	0.0390	0.0465	0.0922	0.1287	0.0464	0.0556	0.0825	0.1180	0.0437	0.0526
DuoRec	0.0667	0.0962	0.0407	0.0481	0.0863	0.1272	0.0432	0.0535	0.0878	0.1244	0.0451	0.0543
FDSA	0.0668	0.0966	0.0403	0.0478	0.0900	0.1327	0.0456	0.0563	0.0839	0.1209	0.0439	0.0532
NOVA	0.0670	0.0952	0.0407	0.0478	0.0951	0.1382	0.0467	0.0575	0.0866	0.1240	0.0441	0.0535
DIF-SR	0.0673	0.0988	0.0412	0.0491	0.0983	0.1419	0.0482	0.0592	0.0871	0.1234	0.0434	0.0526
ASIF	**0.0768**	**0.1131**	**0.0452**	**0.0543**	**0.1131**	**0.1631**	**0.0574**	**0.0700**	**0.0922**	**0.1322**	**0.0453**	**0.0554**
Impr.	14.12%	14.47%	9.71%	10.59%	13.44%	13.98%	15.96%	15.89%	4.77%	6.27%	0.22%	2.03%

Table 2: Overall Performance (HR and NDCG) on public datasets.

Experiments

Table 3: Performance on the industrial dataset.

Model	Industrial			
	H@10	H@20	N@10	N@20
Bert4Rec	0.0706	0.1187	0.0355	0.0476
Caser	0.0808	0.1315	0.0417	0.0544
GRU4Rec	0.0322	0.0575	0.0190	0.0250
SASRec	0.0942	0.1518	<u>0.0480</u>	0.0625
LightSANs	0.0935	0.1556	0.0466	0.0622
FMLP	0.0939	0.1553	0.0454	0.0608
GRU4Rec_F	0.0830	0.1364	0.0433	0.0567
SASRec_F	0.0877	0.1385	0.0463	0.0591
LightSANs_F	0.0889	0.1457	0.0454	0.0596
FMLP_F	0.0863	0.1438	0.0420	0.0564
CL4SRec	0.0683	0.1134	0.0342	0.0455
DuoRec	0.0917	0.1475	0.0475	0.0615
FDSA	0.0913	0.1496	0.0479	<u>0.0626</u>
NOVA	0.0933	0.1517	0.0456	0.0602
DIF-SR	<u>0.0951</u>	<u>0.1559</u>	0.0459	0.0612
ASIF	**0.0996**	**0.1653**	**0.0495**	**0.0660**
Impr.	4.73%	6.03%	3.13%	5.43%

Experiments

Table 4: Ablation results (HR@20 and NDCG@20) on three public datasets. Each row removes a single component from the model except the last row.

Model	Yelp		AliEC		Beauty	
	H@20	N@20	H@20	N@20	H@20	N@20
w/o RSA	0.1075	0.0524	0.1558	0.0668	0.1292	0.0540
w/o HIE	0.0996	0.0493	0.1439	0.0603	0.1255	0.0543
w/o UP	0.1077	0.0522	0.1572	0.0673	0.1298	0.0550
w/o FA	0.1108	0.0534	0.1601	0.0683	0.1317	0.0544
ASIF	**0.1131**	**0.0543**	**0.1631**	**0.0700**	**0.1322**	**0.0554**

Experiments

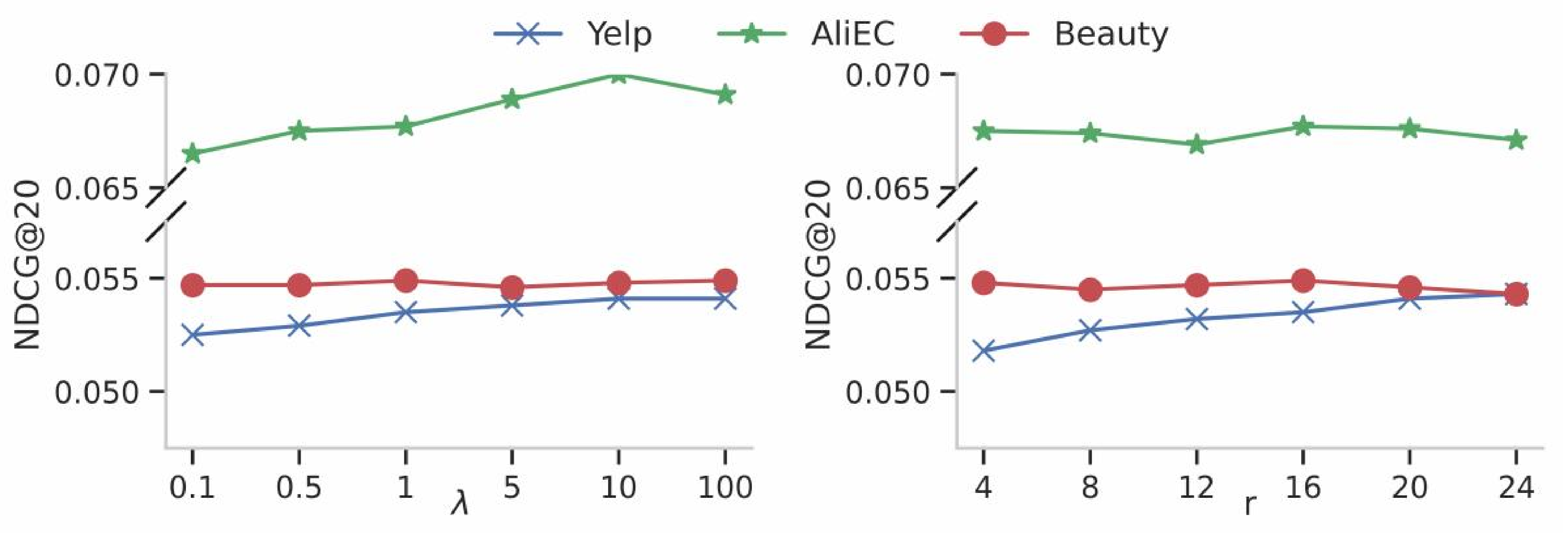


Figure 6: Influence of balance parameter λ and number of orthogonal bases r.

Experiments

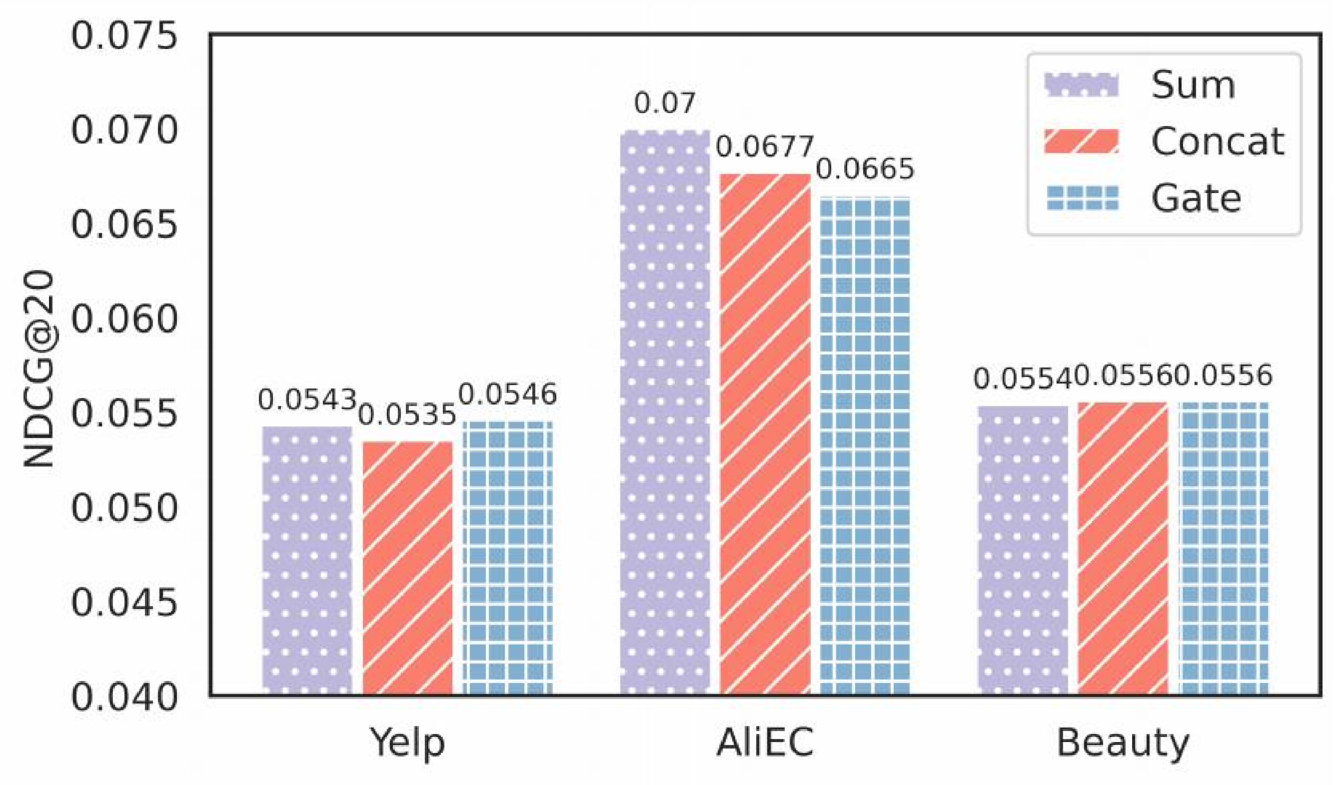


Figure 7: Impact of fusion func $\mathcal{F}$.

Experiments

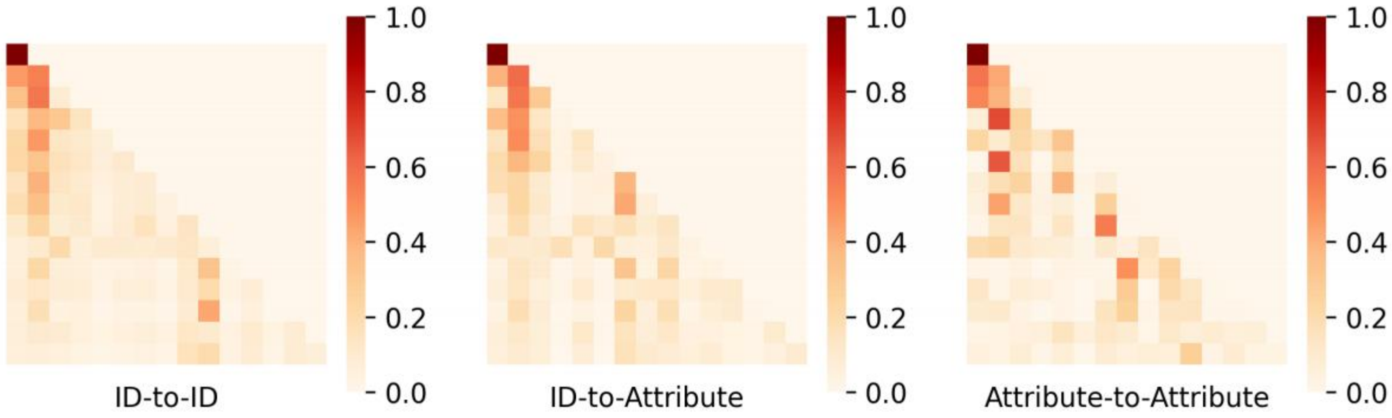


Figure 8: Visualization of attention correlations in ASIF.

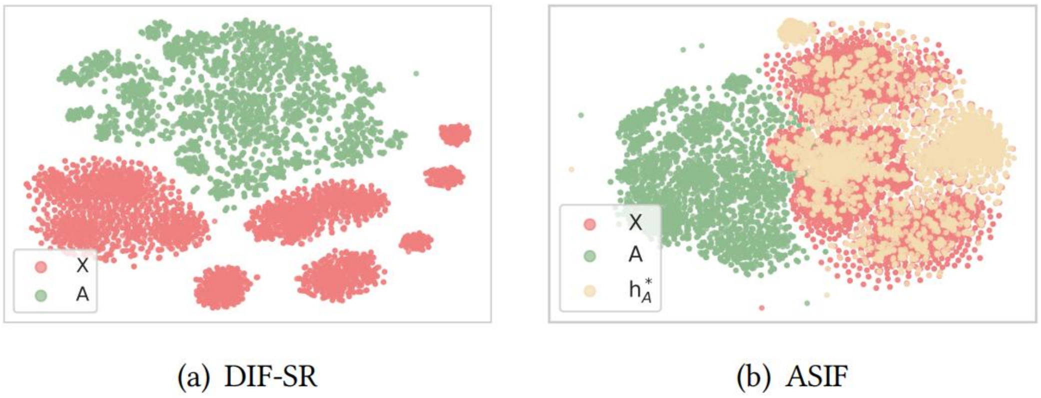


Figure 9: Visualization of clustered embeddings on Yelp.

Experiments

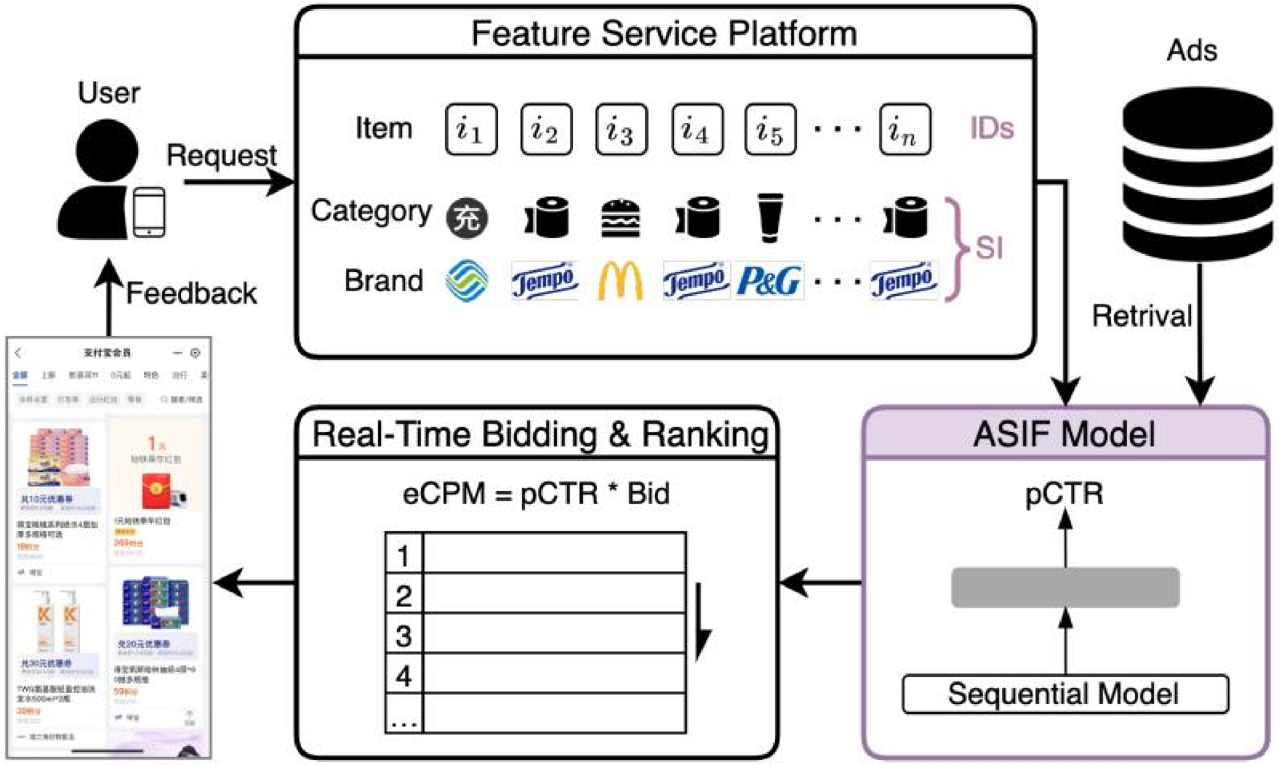


Figure 10: Online deployment of ASIF in Alipay.

Thanks