When Search Meets Recommendation: Learning Disentangled Search Representation for Recommendation

Zihua Si
Gaoling School of Artificial Intelligence
Renmin University of China
Beijing, China
zihua_si@ruc.edu.cn

Zhongxiang Sun
Gaoling School of Artificial Intelligence
Renmin University of China
Beijing, China
sunzhongxiang@ruc.edu.cn

Xiao Zhang*
Jun Xu*
Gaoling School of Artificial Intelligence
Renmin University of China
Beijing, China
{zhangx89,junxu}@ruc.edu.cn

Xiaoxue Zang
Yang Song
Kuaishou Technology Co., Ltd.
Beijing, China
{zangxiaoxue,yangsong}@kuaishou.com

Kun Gai
Unaffiliated
Beijing, China
gai.kun@qq.com

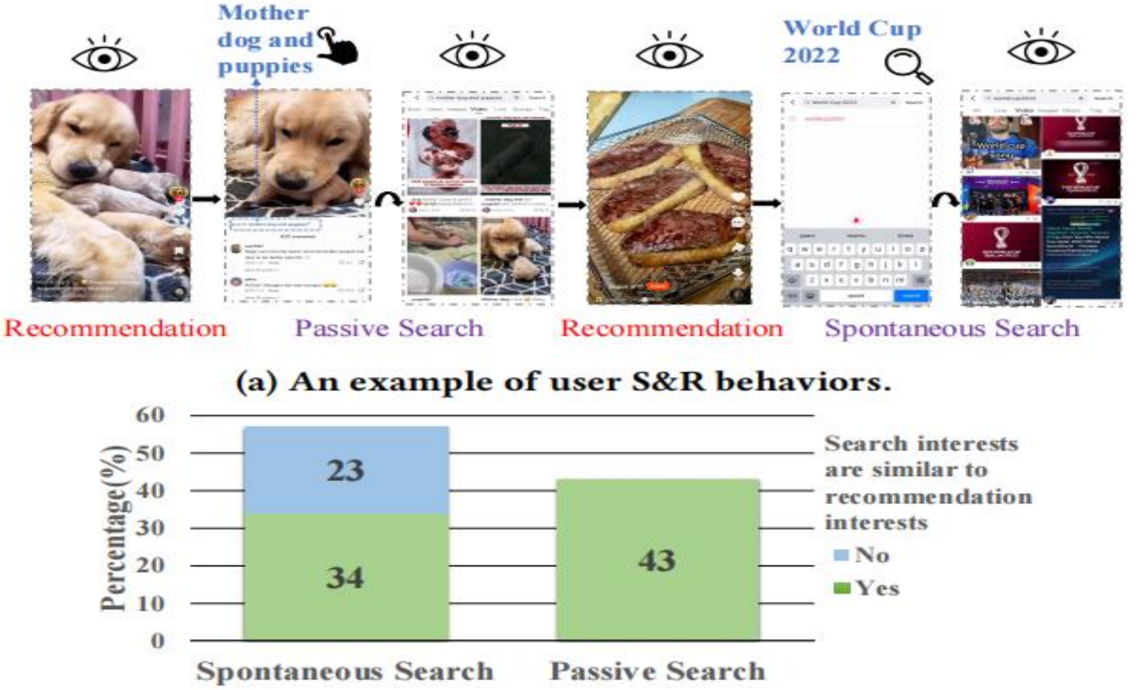
Ji-Rong Wen
Gaoling School of Artificial Intelligence
Renmin University of China
Beijing, China
jrwen@ruc.edu.cn

SIGIR 2023

Code&data： https://github.com/Ethan00Si/SESREC-SIGIR-2023

Reported by Changjiang Hu

Introduction



(b) Distribution of spontaneous and passive search w.r.t. recommendation interests.

Figure 1: S&R behaviors in the short-video scenario. (a) After watching a video about dogs, the user chooses to click on the suggested query (passive search) to explore more. Later, after watching a food video, the user searches “world cup 2022”, a spontaneous search unrelated to the watched video. (b) Statistics of search behaviors collected from the Kuaishou app. 57% of the search behaviors are spontaneous and 43% are passive. 23% of the spontaneous searches have dissimilar interests to the recommendation interests.

Method

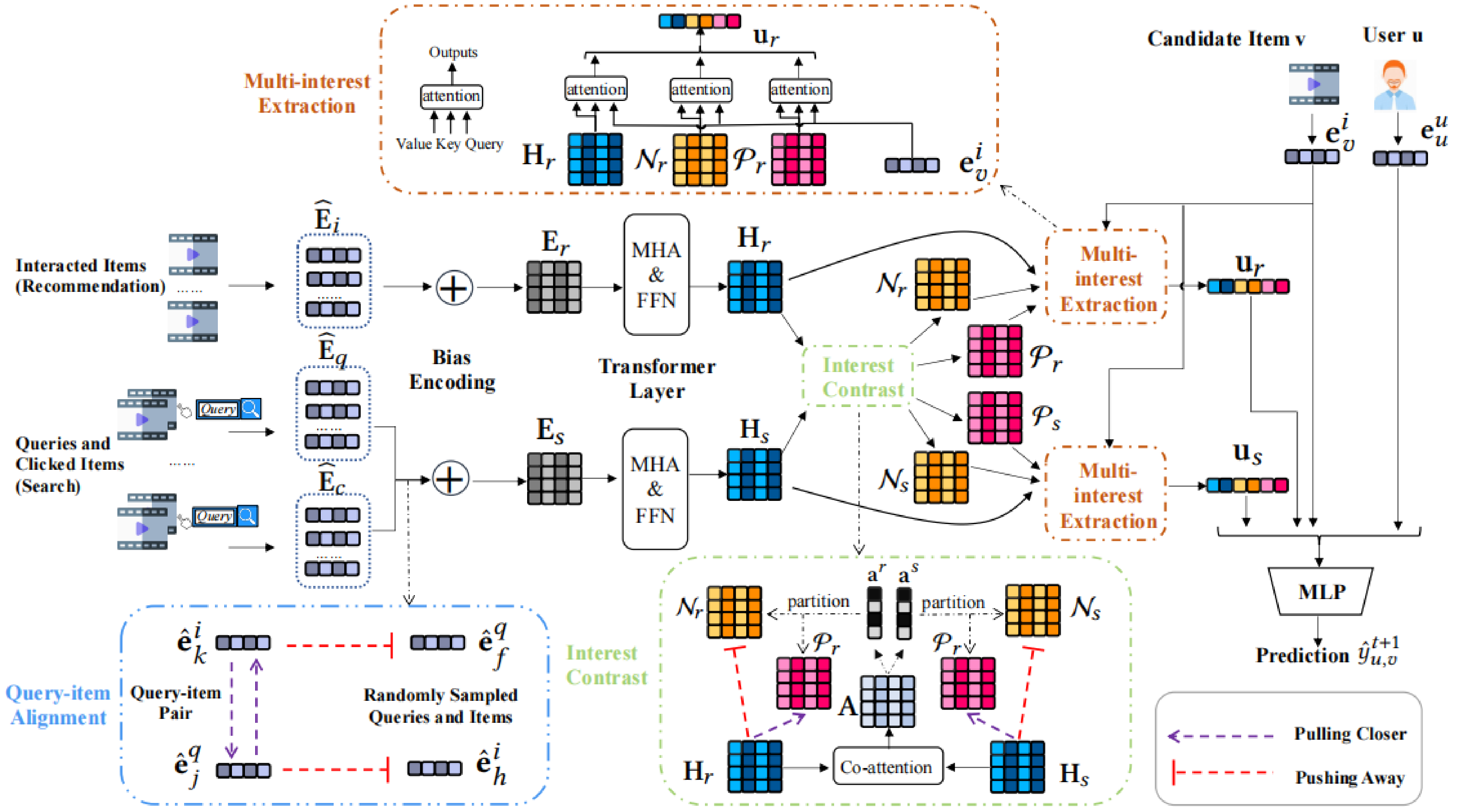


Figure 2: The architecture of SESRec. From left to right is the process of modeling S&R histories. On the far right is the process of ultimate prediction. The three colored modules with dashed lines conduct interest disentanglement.

Method

PROBLEM FORMULATION:

user: $\mathcal{U}$ **item:** $\mathcal{I}$ **query:** $\mathcal{Q}$

$$S_i^u = [i_1, i_2, \ldots, i_{T_r}]$$

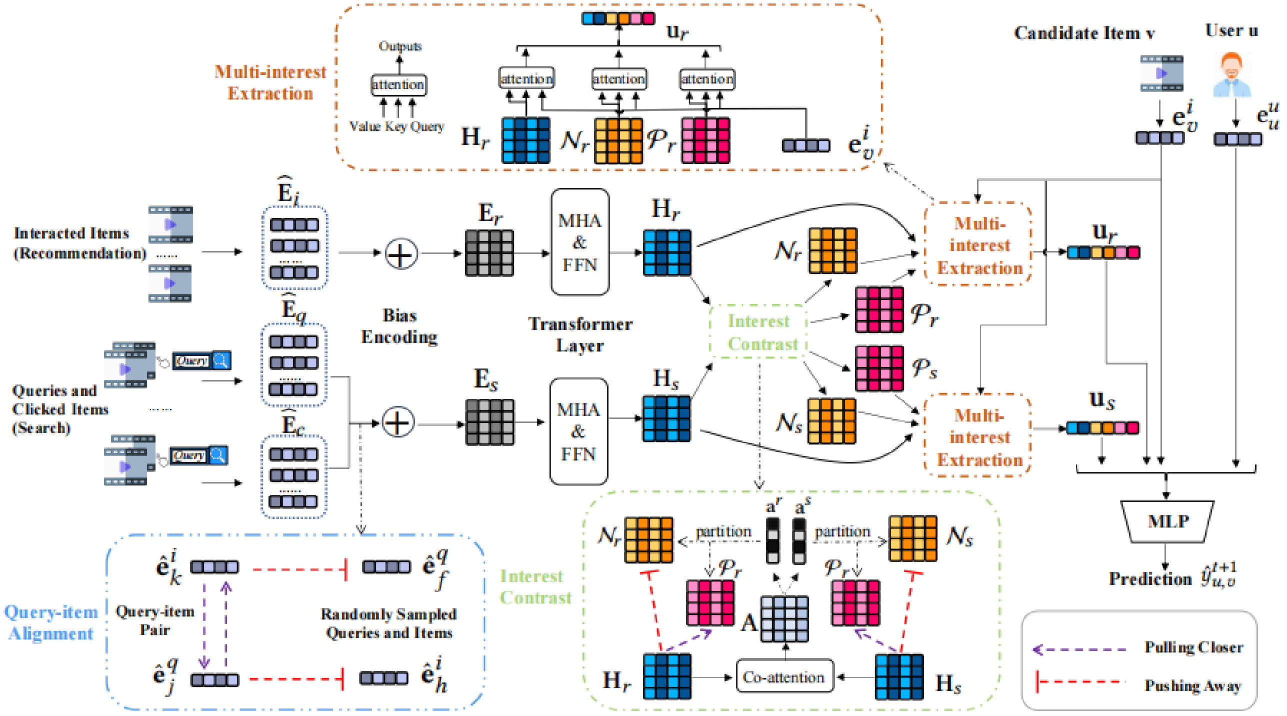
$$S_q^u = [q_1, q_2, \ldots, q_{T_s}]$$

$$S_c^u = \left[i_{q_1}^{(1)}, i_{q_1}^{(2)}, i_{q_2}^{(1)}, \ldots, i_{T_s}^{(1)}, i_{T_s}^{(2)}, i_{T_s}^{(3)}\right]$$

$$P(i_{t+1} \mid S_i^u, S_q^u, S_c^u)$$

Method

Encoding Sequential Behaviors



Embedding Layer

$$\mathbf{e}^u = \mathbf{e}^{\mathrm{ID}_u} \| \mathbf{e}^{a_1} \| \cdots \| \mathbf{e}^{a_n} \quad (\mathbf{e}^i = \mathbf{e}^{\mathrm{ID}_i} \| \mathbf{e}^{b_1} \| \cdots \| \mathbf{e}^{b_m})$$

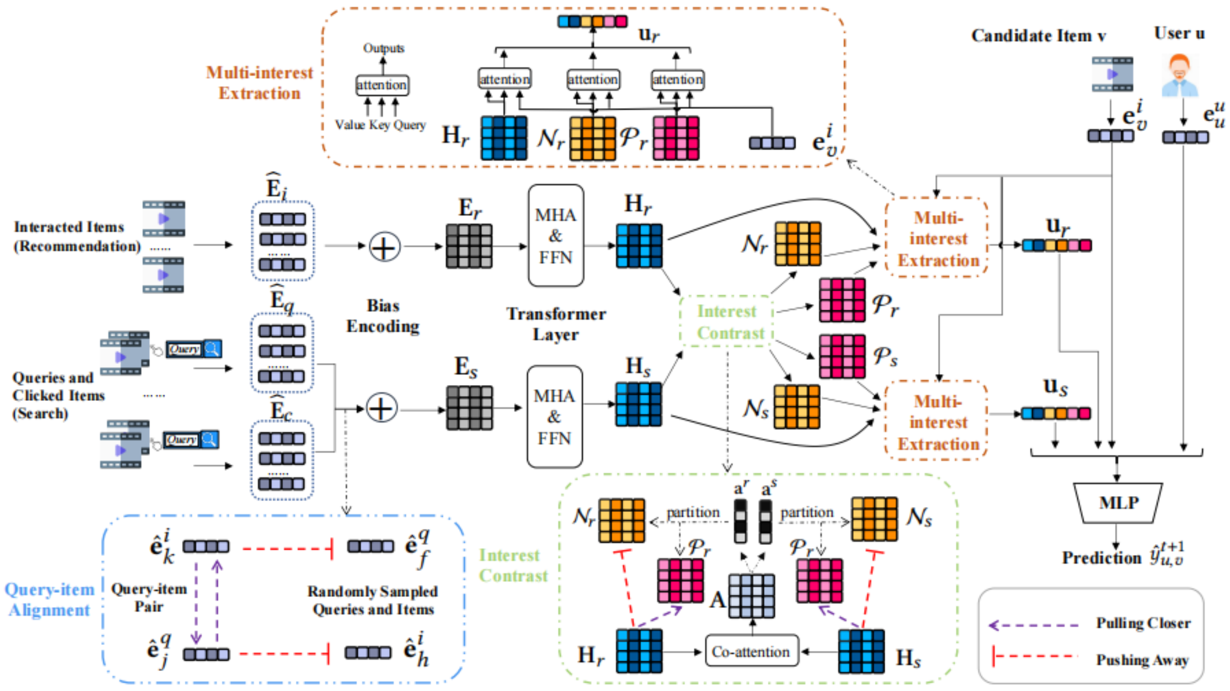
$$\mathbf{e}^q = \mathbf{e}^{\mathrm{ID}_q} \| \mathrm{MEAN}(\mathbf{e}^{w_1}, \mathbf{e}^{w_2}, \ldots, \mathbf{e}^{w_{|q|}})$$

$$S_i^u \quad \mathbf{E}_i = [\mathbf{e}_1^i, \mathbf{e}_2^i, \ldots, \mathbf{e}_{T_r}^i]^\top \in \mathbb{R}^{T_r \times d_i}$$

$$S_q^u \quad \mathbf{E}_q = [\mathbf{e}_1^q, \mathbf{e}_2^q, \ldots, \mathbf{e}_{T_s}^q]^\top \in \mathbb{R}^{T_s \times d_q}$$

$$S_c^u \quad \mathbf{E}_c = [\mathbf{e}_1^i, \mathbf{e}_2^i, \ldots, \mathbf{e}_{|S_c^u|}^i]^\top \in \mathbb{R}^{|S_c^u| \times d_i}$$

$$\widehat{\mathbf{E}}_i = \mathbf{E}_i \mathbf{W}_i, \quad \widehat{\mathbf{E}}_q = \mathbf{E}_q \mathbf{W}_q, \quad \widehat{\mathbf{E}}_c = \mathbf{E}_c \mathbf{W}_i, \qquad (1)$$



Bias Encoding And Query-item Alignment

$$\mathbf{E}_r = \widehat{\mathbf{E}}_i + \mathbf{P}_r \qquad (2)$$

$$\mathbf{E}_q = [\mathbf{e}_1^q, \mathbf{e}_2^q, \ldots, \mathbf{e}_{T_s}^q]^\top \in \mathbb{R}^{T_s \times d_q} \qquad S_q^u = [q_1, q_2, \ldots, q_{T_s}]$$

$$\mathbf{E}_c = [\mathbf{e}_1^i, \mathbf{e}_2^i, \ldots, \mathbf{e}_{|S_c^u|}^i]^\top \in \mathbb{R}^{|S_c^u| \times d_i} \qquad S_c^u = \left[i_{q_1}^{(1)}, i_{q_1}^{(2)}, i_{q_2}^{(1)}, \ldots, i_{T_s}^{(1)}, i_{T_s}^{(2)}, i_{T_s}^{(3)}\right]$$

query-to-item alignment

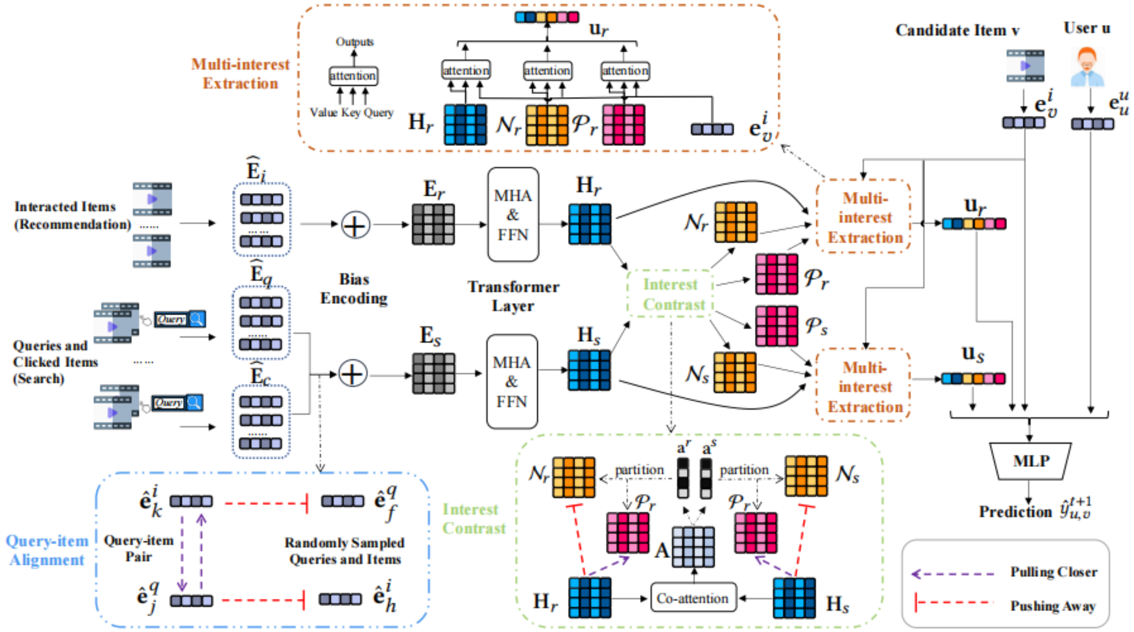
$$\mathcal{L}_{\mathrm{A_{q2i}}}^{u,t} = -\sum_{j=1}^{T_s} \sum_{k=1}^{|q_j|} \log \frac{\exp(s(\hat{\mathbf{e}}_j^q, \hat{\mathbf{e}}_k^i)/\tau)}{\sum_{h \in \mathcal{I}_{\mathrm{neg}}} \exp(s(\hat{\mathbf{e}}_j^q, \hat{\mathbf{e}}_h^i)/\tau)}, \qquad (6)$$

item-to-query alignment

$$\mathcal{L}_{\mathrm{A_{i2q}}}^{u,t} = -\sum_{j=1}^{T_s} \sum_{k=1}^{|q_j|} \log \frac{\exp(s(\hat{\mathbf{e}}_j^q, \hat{\mathbf{e}}_k^i)/\tau)}{\sum_{f \in \mathcal{Q}_{\mathrm{neg}}} \exp(s(\hat{\mathbf{e}}_f^q, \hat{\mathbf{e}}_k^i)/\tau)}, \qquad (7)$$

$$\mathcal{L}_{\mathrm{ali}}^{u,t} = \frac{1}{2}(\mathcal{L}_{\mathrm{A_{q2i}}}^{u,t} + \mathcal{L}_{\mathrm{A_{i2q}}}^{u,t}), \qquad (8)$$

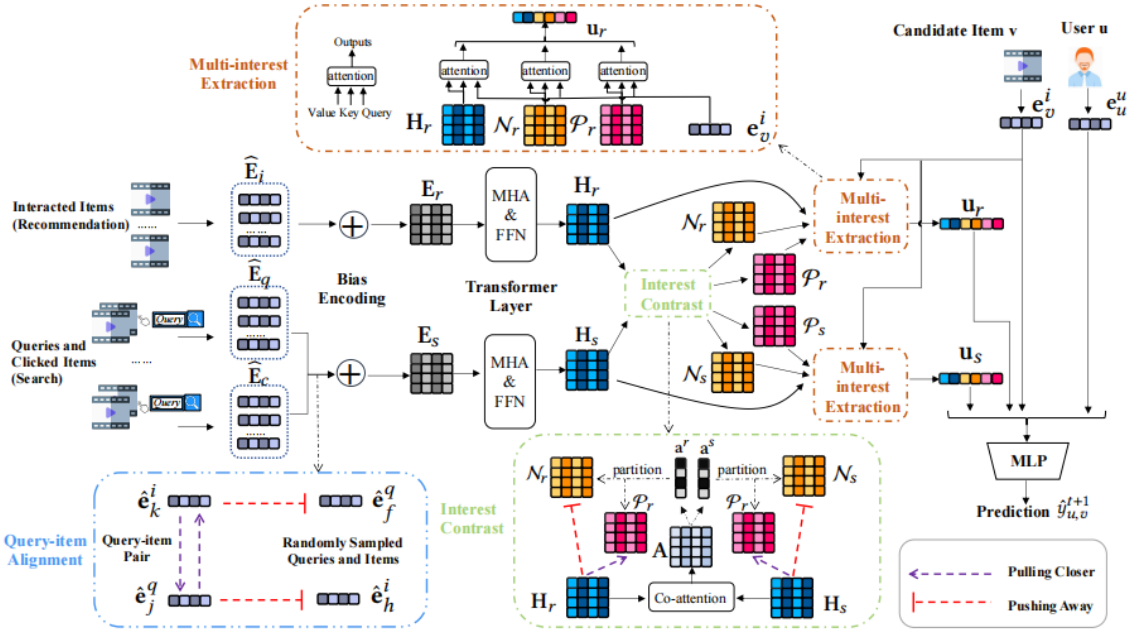
$$\mathbf{E}_s = \widehat{\mathbf{E}}_q + \widetilde{\mathbf{E}}_c + \mathbf{P}_s + \widehat{\mathbf{M}}_s \qquad (3)$$



Transformer Layer

$$\mathbf{F}_s = \mathrm{MHA}_s(\mathbf{E}_s), \quad \mathbf{F}_r = \mathrm{MHA}_r(\mathbf{E}_r), \tag{4}$$

$$\mathbf{H}_s = \mathrm{FFN}_s(\mathbf{F}_s), \quad \mathbf{H}_r = \mathrm{FFN}_r(\mathbf{F}_r), \tag{5}$$



Interest Contrast

$$\mathbf{A} = \tanh(\mathbf{H}_s \mathbf{W}_l (\mathbf{H}_r)^{\mathrm{T}}), \tag{9}$$

$$\mathbf{a}^s = \mathrm{softmax}(\mathbf{W}_r \mathbf{H}_r^{\mathrm{T}} \mathbf{A}^{\mathrm{T}}), \quad \mathbf{a}^r = \mathrm{softmax}(\mathbf{W}_s \mathbf{H}_s^{\mathrm{T}} \mathbf{A}), \tag{10}$$

$$\mathcal{P}_s = \{\mathbf{h}_j^s \mid \mathbf{a}_j^s > \gamma_s\}, \quad \mathcal{N}_s = \{\mathbf{h}_j^s \mid \mathbf{a}_j^s \le \gamma_s\}, \tag{11}$$

$$\mathcal{P}_r = \{\mathbf{h}_j^r \mid \mathbf{a}_j^r > \gamma_r\}, \quad \mathcal{N}_r = \{\mathbf{h}_j^r \mid \mathbf{a}_j^r \le \gamma_r\}, \tag{12}$$

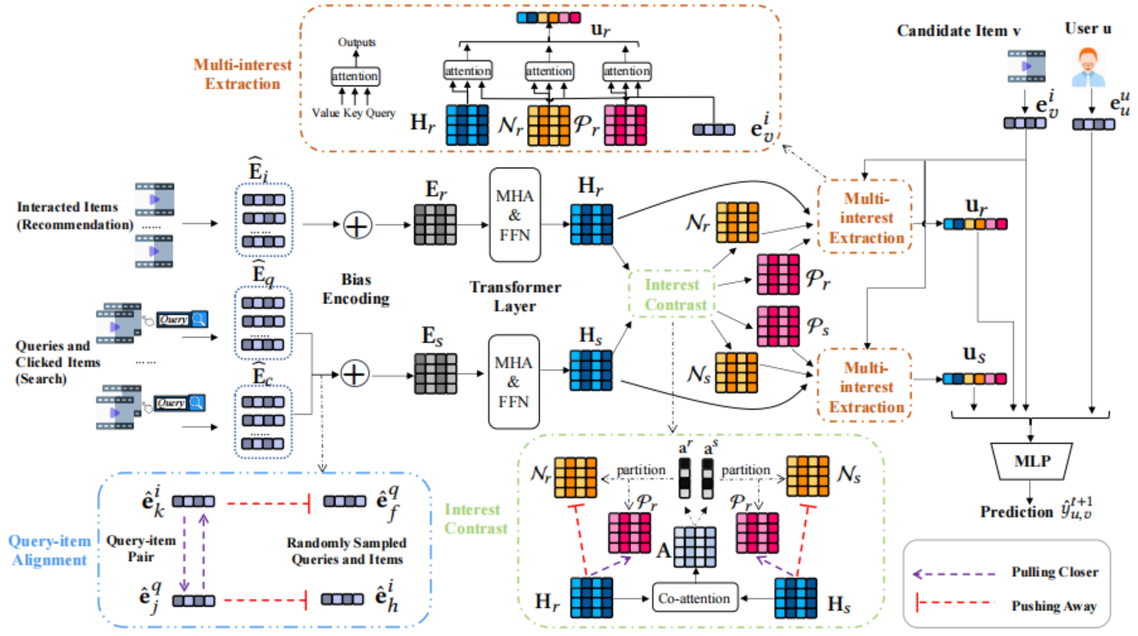
$$\mathbf{i}_s^A = \sum_{j=1}^{T_s} \mathbf{a}_j^s \mathbf{h}_j^s, \quad \mathbf{i}_s^P = \mathrm{MEAN}(\mathcal{P}_s), \quad \mathbf{i}_s^N = \mathrm{MEAN}(\mathcal{N}_s), \tag{13}$$

$$\mathbf{i}_r^A = \sum_{j=1}^{T_r} \mathbf{a}_j^r \mathbf{h}_j^r, \quad \mathbf{i}_r^P = \mathrm{MEAN}(\mathcal{P}_r), \quad \mathbf{i}_r^N = \mathrm{MEAN}(\mathcal{N}_r), \tag{14}$$

$$\mathcal{L}_{\mathrm{tri}}(a, p, n) = \max\{d(a, p) - d(a, n) + m, 0\}, \tag{15}$$

$$\mathcal{L}_{\mathrm{con}}^{u,t} = \mathcal{L}_{\mathrm{tri}}(\mathbf{i}_r^A, \mathbf{i}_r^P, \mathbf{i}_r^N) + \mathcal{L}_{\mathrm{tri}}(\mathbf{i}_s^A, \mathbf{i}_s^P, \mathbf{i}_s^N), \tag{16}$$

Multi-interest Extraction

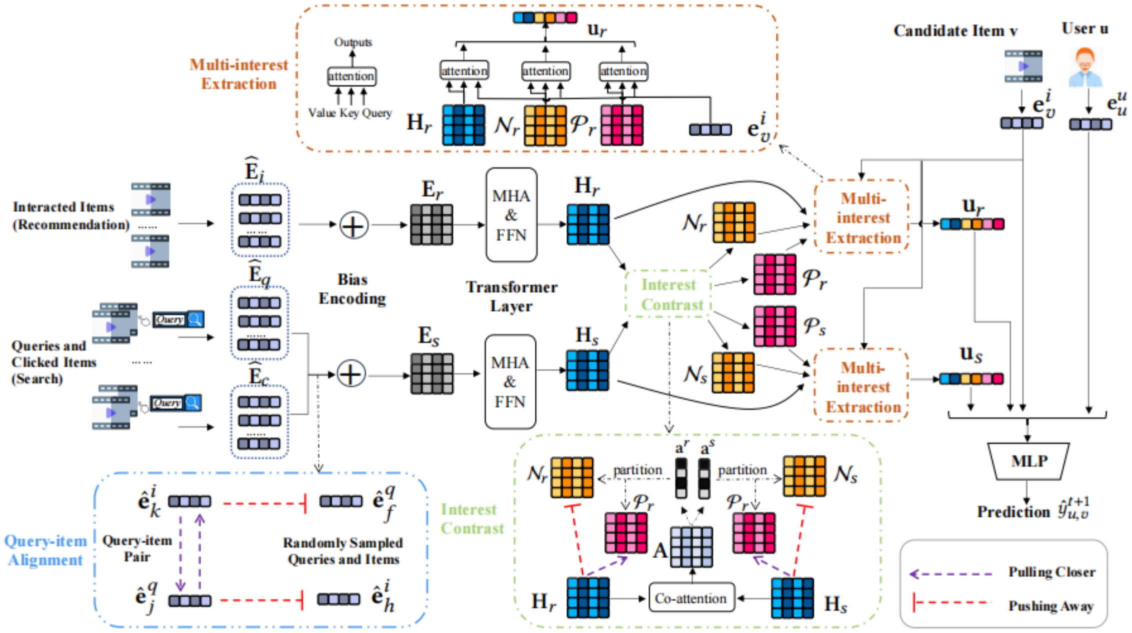


$$\mathbf{u}_r^{\text{all}} = \sum_{j=1}^{T_r} a_j^{\text{all}} \mathbf{h}_j^r, \quad a_j^{\text{all}} = \frac{\exp((\mathbf{h}_j^r)^{\mathrm{T}} \mathbf{W}_d \mathbf{e}_v^i)}{\sum_{k=1}^{T_r} \exp((\mathbf{h}_k^r)^{\mathrm{T}} \mathbf{W}_d \mathbf{e}_v^i)}, \tag{17}$$

$$\mathbf{u}_r^{\text{sim}} = \sum_{\mathbf{h}_j^r \in \mathcal{P}_r} a_j^{\text{sim}} \mathbf{h}_j^r, \quad a_j^{\text{sim}} = \frac{\exp((\mathbf{h}_j^r)^{\mathrm{T}} \mathbf{W}_d \mathbf{e}_v^i)}{\sum_{\mathbf{h}_k^r \in \mathcal{P}_r} \exp((\mathbf{h}_k^r)^{\mathrm{T}} \mathbf{W}_d \mathbf{e}_v^i)}, \tag{18}$$

$$\mathbf{u}_r^{\text{diff}} = \sum_{\mathbf{h}_j^r \in \mathcal{N}_r} a_j^{\text{diff}} \mathbf{h}_j^r, \quad a_j^{\text{diff}} = \frac{\exp((\mathbf{h}_j^r)^{\mathrm{T}} \mathbf{W}_d \mathbf{e}_v^i)}{\sum_{\mathbf{h}_k^r \in \mathcal{N}_r} \exp((\mathbf{h}_k^r)^{\mathrm{T}} \mathbf{W}_d \mathbf{e}_v^i)}, \tag{19}$$

$$\mathbf{u}_r = \mathbf{u}_r^{\text{all}} \| \mathbf{u}_r^{\text{sim}} \| \mathbf{u}_r^{\text{diff}}, \tag{20}$$



Prediction and Model Training

$$\hat{y}_{u,v}^{t+1} = \mathrm{MLP}(\mathbf{u}_r \| \mathbf{u}_s \| \mathbf{e}_v^i \| \mathbf{e}_u^u), \qquad (21)$$

$$\mathcal{L}_{\mathrm{rec}}^{u,t} = -\frac{1}{N} \sum_{v \in O} y_{u,v}^{t+1} \log(\hat{y}_{u,v}^{t+1}) + (1 - y_{u,v}^{t+1}) \log(1 - \hat{y}_{u,v}^{t+1}), \qquad (22)$$

$$\mathcal{L} = \sum_{u=1}^{|\mathcal{U}|} \sum_{t=1}^{T_u} (\mathcal{L}_{\mathrm{rec}}^{u,t} + \alpha \mathcal{L}_{\mathrm{ali}}^{u,t} + \beta \mathcal{L}_{\mathrm{con}}^{u,t}) + \lambda ||\Theta||_2. \qquad (23)$$

Experiments

Table 1: Statistics of datasets used in this paper. 'S' and 'R' denote search and recommendation, respectively.

Dataset	Users	Items	Queries	Actions-S	Actions-R
Kuaishou	35,721	822,832	398,924	922,531	11,381,172
Amazon	68,223	61,934	4,298	934,664	989,618

Experiments

Table 2: Overall performance comparisons on both datasets. The best and the second-best performance methods are denote in bold and underlined fonts respectively. * means improvements over the second-best methods are significant (p-value < 0.01)

Dataset		Kuaishou						Amazon (Kindle Store)					
Category	Method	NDCG@5	NDCG@10	HIT@1	HIT@5	HIT@10	MRR	NDCG@5	NDCG@10	HIT@1	HIT@5	HIT@10	MRR
Sequential	STAMP	0.2544	0.2981	0.1413	0.3616	0.4970	0.2569	0.2612	0.3103	0.1336	0.3833	0.5352	0.2608
	DIN	0.2969	0.3418	0.1792	0.4092	0.5484	0.2976	0.2999	0.3495	0.1591	0.4340	0.5871	0.2942
	GRU4Rec	0.3247	0.3688	0.1890	0.4517	0.5881	0.3180	0.3099	0.3662	0.1479	0.4648	0.6388	0.2993
	SASRec	0.3252	0.3693	0.1904	0.4501	0.5864	0.3187	0.3822	0.4312	0.2187	0.5324	0.6838	0.3675
	DIEN	0.3217	0.3704	0.1914	0.4463	0.5969	0.3192	0.3336	0.3803	0.1871	0.4706	0.6150	0.3246
	FMLP-Rec	0.3354	0.3787	0.1953	0.4651	0.5988	0.3270	0.4073	0.4550	0.2349	0.5651	**0.7121**	0.3883
Search-aware	NRHUB	0.2964	0.3431	0.1665	0.4199	0.5647	0.2933	0.2744	0.3265	0.1329	0.4099	0.5708	0.2704
	JSR	0.3015	0.3513	0.1738	0.4241	0.5783	0.3004	0.3221	0.3722	0.2057	0.4386	0.5937	0.3224
	IV4REC	0.3114	0.3591	0.1877	0.4282	0.5761	0.3116	0.3473	0.3960	0.1853	0.4985	0.6258	0.3331
	Query-SeqRec	0.3117	0.3581	0.1740	0.4412	0.5844	0.3055	0.3692	0.4142	0.2187	0.5083	0.6470	0.3572
	SRJGraph	0.3297	0.3762	0.2046	0.4479	0.5917	0.3277	0.3670	0.4043	**0.2760**	0.4898	0.6242	0.3708
	SESRec	**0.3541***	**0.4054***	**0.2173***	**0.4848***	**0.6436***	**0.3490***	**0.4224***	**0.4663***	0.2580	**0.5723***	0.7074	**0.4046***

Experiments

Table 3: Ablation studies by progressively adding proposed modules to the base model. MIE is short for the multi-interest extraction module.

Model	N@5	N@10	H@1	H@5	H@10	MRR
Base	0.3394	0.3770	0.2027	0.4618	0.6094	0.3294
$+\mathcal{L}_{\text{ali}}^{u,t}$	0.3464	0.3982	0.2106	0.4762	0.6308	0.3421
$+\mathcal{L}_{\text{con}}^{u,t}$	0.3507	0.4021	0.2139	0.4812	0.6406	0.3459
+MIE	**0.3541**	**0.4054**	**0.2173**	**0.4848**	**0.6436**	**0.3490**

Experiments

Table 4: The analysis of the positive/negative selection thresholds γ_r, γ_s in interest disentanglement, as defined in Equation (11) and (12).

γ_r, γ_s	N@5	N@10	H@1	H@5	H@10	MRR
1/16	0.3429	0.3927	0.2125	0.4681	0.6224	0.3429
1/8	0.3449	0.3940	0.2148	0.4691	0.6211	0.3415
Median	0.3510	0.4035	0.2135	0.4828	**0.6453**	0.3462
Mean	**0.3541**	**0.4054**	**0.2173**	**0.4848**	0.6436	**0.3490**

Experiments

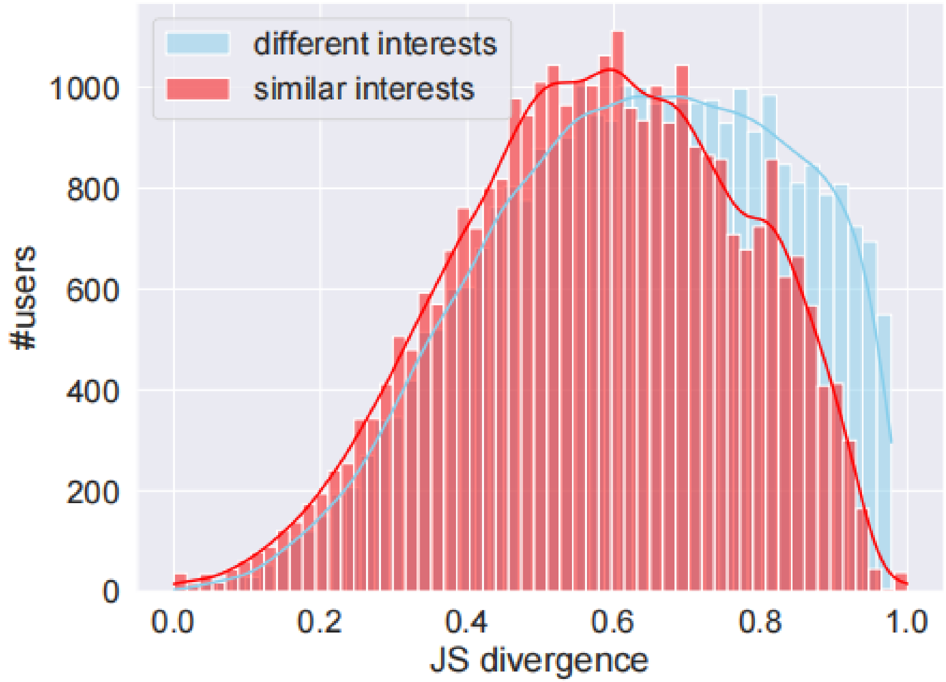


Figure 3: Visualization of similarity between similar and dissimilar interests in S&R behaviors for all users based on a histogram. We use the JS divergence of item category distributions to estimate similarities between S&R behaviors. Similar interests are more similar than dissimilar interests with smaller values of JS divergence.

Experiments

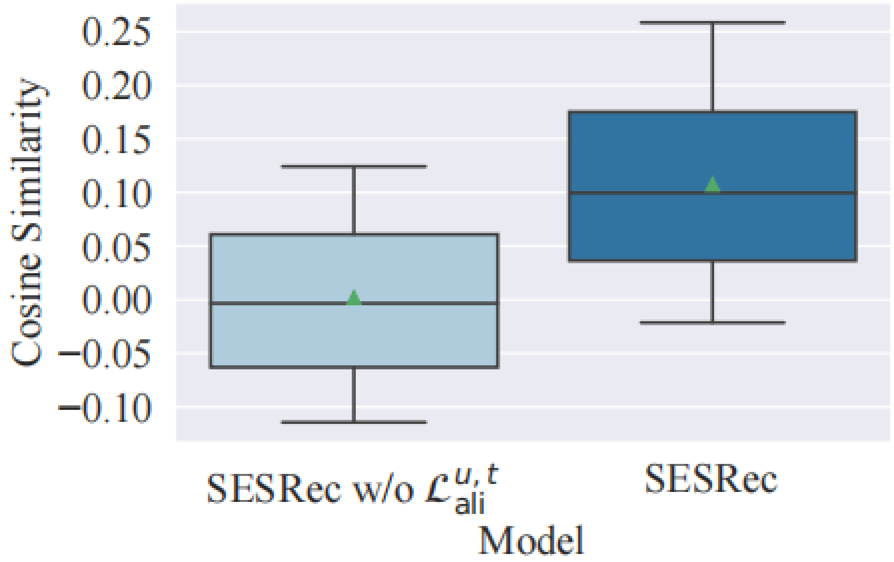
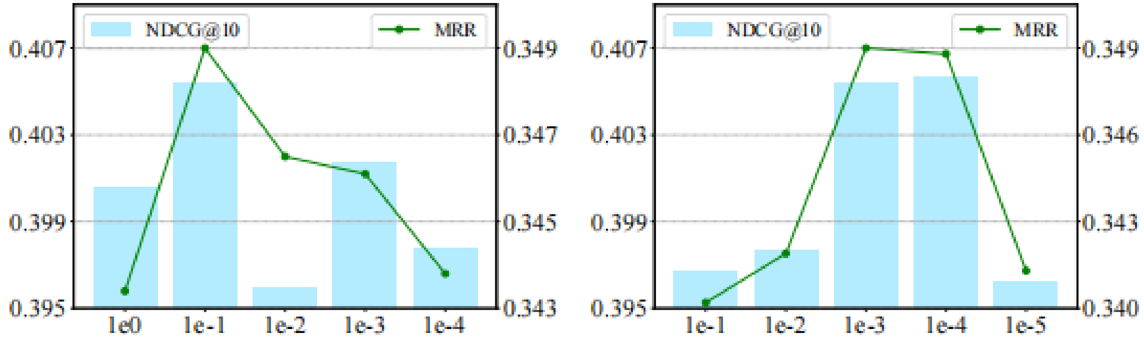


Figure 4: Distribution of cosine similarity between representations of queries and corresponding items based on box plots. Rectangles denote mean values. With query-item alignment loss $\mathcal{L}_{\text{ali}}^{u,t}$, embeddings of query-item pairs are more similar with higher cosine values.

Experiments



(a) Performance of different α. (b) Performance of different β.

Figure 5: Effects of hyper-parameters α and β in terms of NDCG@10 and MRR.

Thanks