Dual-Task Learning for Multi-Behavior Sequential Recommendation

Jinwei Luo
Shenzhen University
Shenzhen, China
luojinwei2016@email.szu.edu.cn

Mingkai He
Xiaolin Lin
Shenzhen University
Shenzhen, China

Weike Pan*
Shenzhen University
Shenzhen, China
panweike@szu.edu.cn

Zhong Ming*
Shenzhen University
Shenzhen, China
mingz@szu.edu.cn

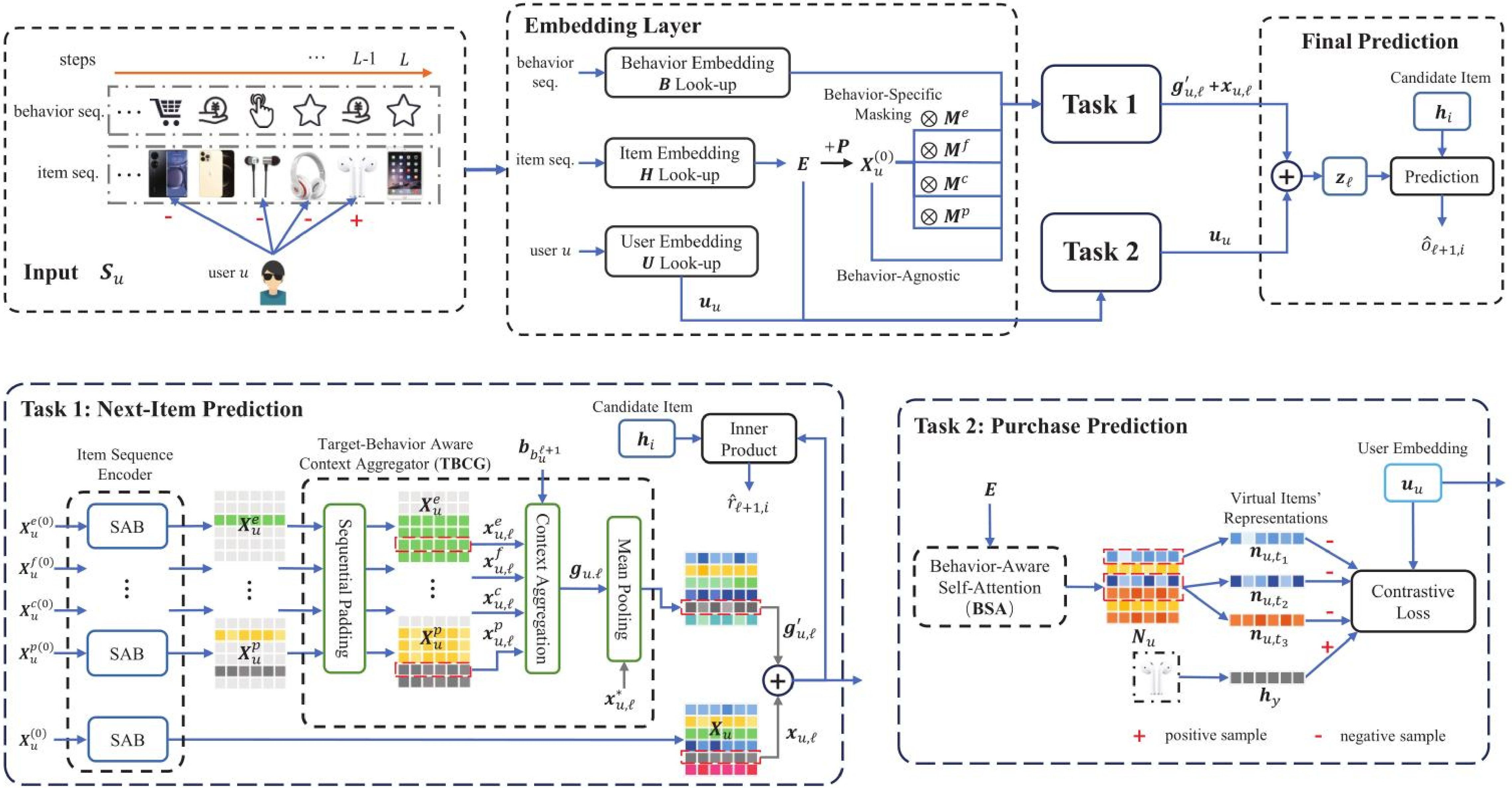
CIKM2022

Reported by Minqin Li

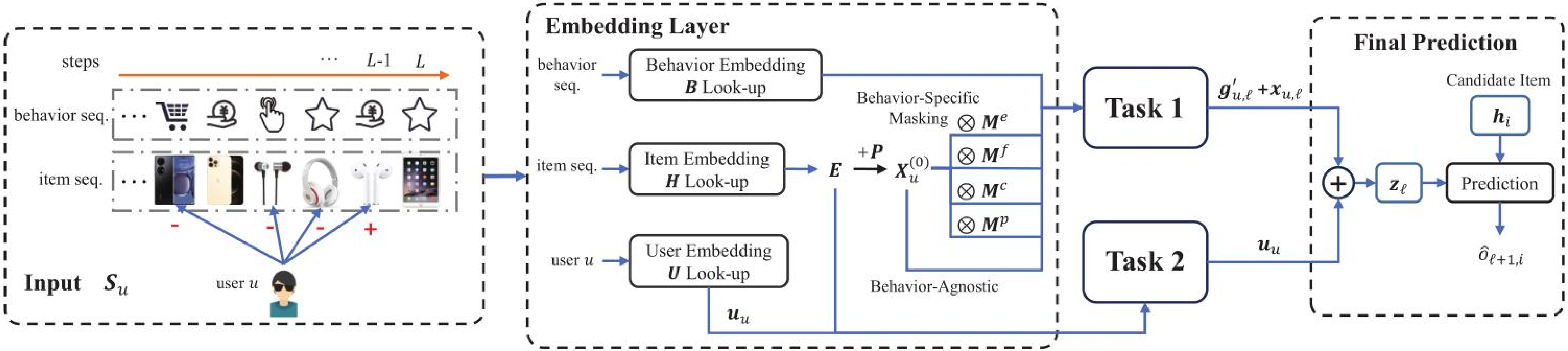
Introduction

- MBSR that exploits users' heterogeneous interactions in sequences has received relatively little attention.
- Existing works often overlook the complementary effect of different perspectives when addressing the MBSR problem.
  - Some works take a multi-behavior interaction sequence as a behavior-agnostic item sequence and a behavior sequence.
  - Some works take a multi-behavior interaction sequence as some behavior-specific item sub-sequences.

Method

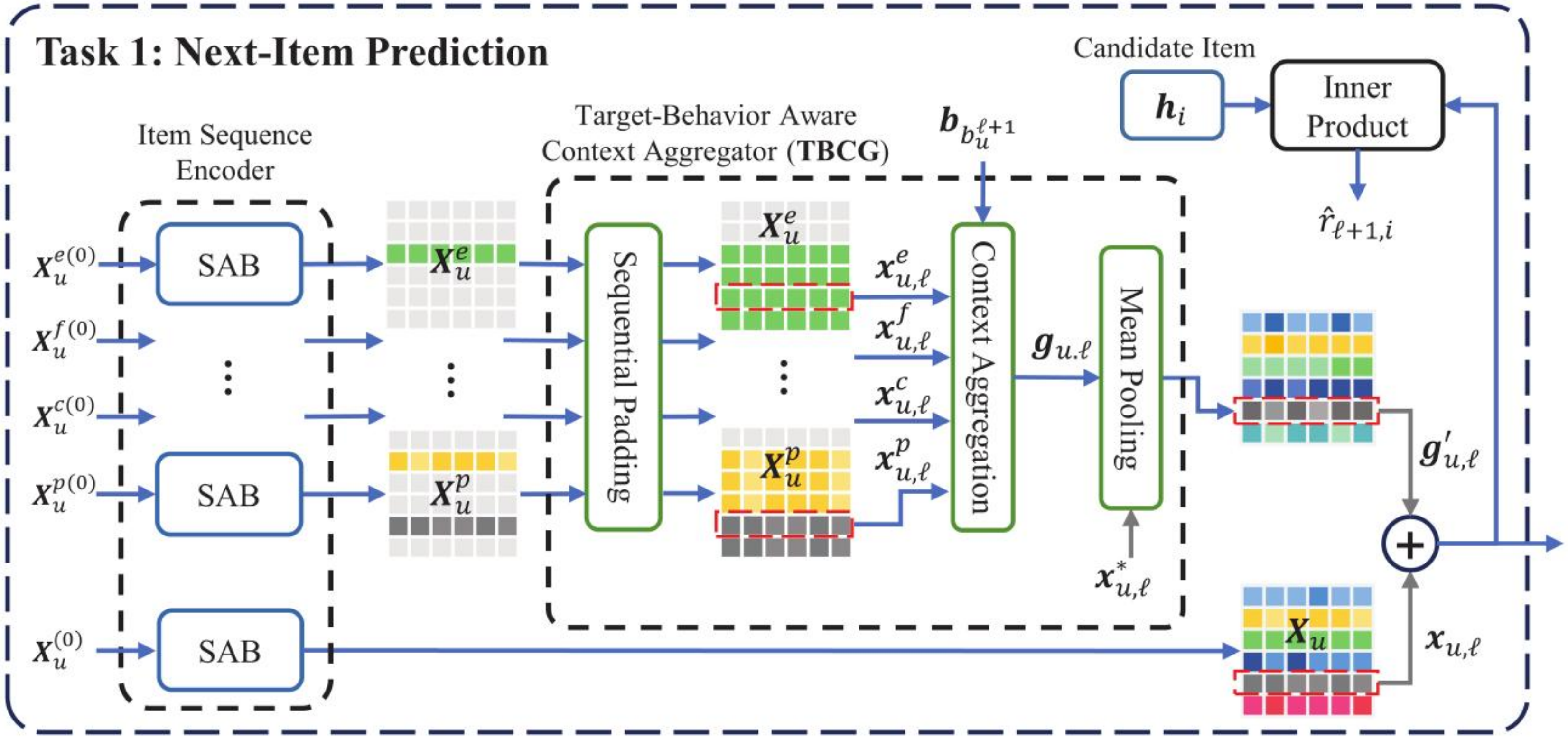

Input S_u
steps
⋯ L-1 L
behavior seq.
item seq.
user u
Embedding Layer
Behavior Embedding
B Look-up
Item Embedding
H Look-up
User Embedding
U Look-up
Behavior-Specific Masking
Behavior-Agnostic
Task 1
Task 2
Final Prediction
Candidate Item
Prediction
Task 1: Next-Item Prediction
Item Sequence Encoder
SAB
Target-Behavior Aware Context Aggregator (TBCG)
Sequential Padding
Context Aggregation
Mean Pooling
Candidate Item
Inner Product
Task 2: Purchase Prediction
User Embedding
Behavior-Aware Self-Attention (BSA)
Virtual Items' Representations
Contrastive Loss
positive sample
negative sample

Method



$$X_u^{e(0)} = X_u^{(0)} \otimes M^e \qquad (1)$$

Method



$$Q'_\ell = b_{b_u^{\ell+1}} W_{Q'} \quad (5)$$

$$K'_\ell = [x^e_{u,\ell}; x^f_{u,\ell}; x^c_{u,\ell}; x^p_{u,\ell}] W_{K'} \quad (6)$$

$$V'_\ell = [x^e_{u,\ell}; x^f_{u,\ell}; x^c_{u,\ell}; x^p_{u,\ell}] W_{V'} \quad (7)$$

$$g_{u,\ell} = softmax(\frac{Q'_\ell K'^T_\ell}{\sqrt{d}}) V'_\ell \quad (8)$$

$$g'_{u,\ell} = \frac{g_{u,\ell} + x^*_{u,\ell}}{2} \quad (9)$$

$$SAB(X) = FFL(SAL(X)) \quad (2)$$

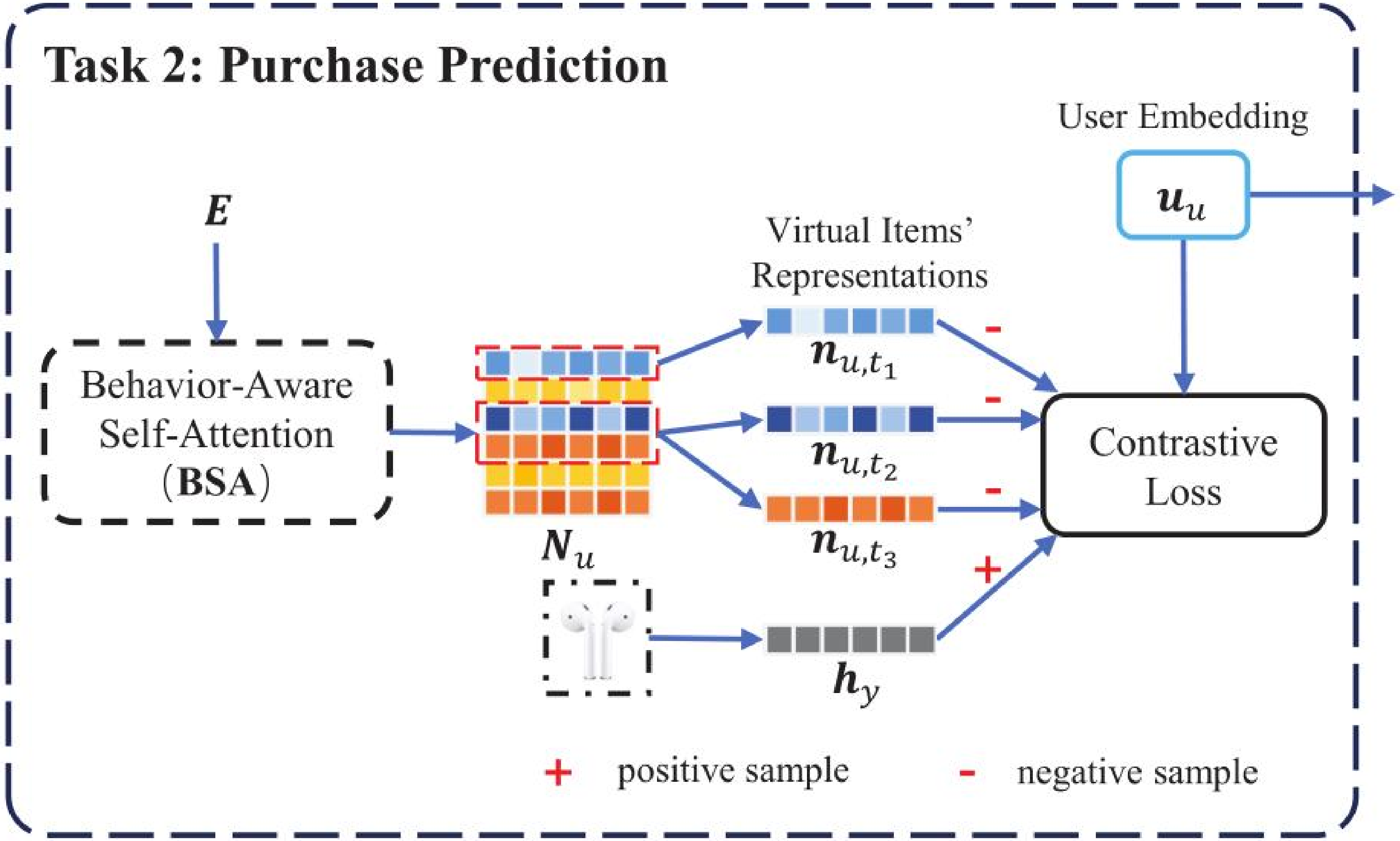
$$X' = SAL(X) = (softmax(\frac{QK^T}{\sqrt{d}}) \otimes \Delta) V \quad (3)$$

$$FFL(X') = ReLU(X'W_1 + b_1) W_2 + b_2 \quad (4)$$

$$\hat{r}_{\ell+1,i} = (x_{u,\ell} + g'_{u,\ell})(h_i)^T \quad (10)$$

$$\mathcal{L}_1 = -\sum_{u \in \mathcal{U}} \sum_{\ell=2}^{L+1} \delta(i_u^\ell)[\log(\sigma(\hat{r}_{\ell, i_u^\ell})) + \log(1 - \sigma(\hat{r}_{\ell,j}))] \quad (11)$$

Method



$$A = M \otimes \Delta \tag{12}$$

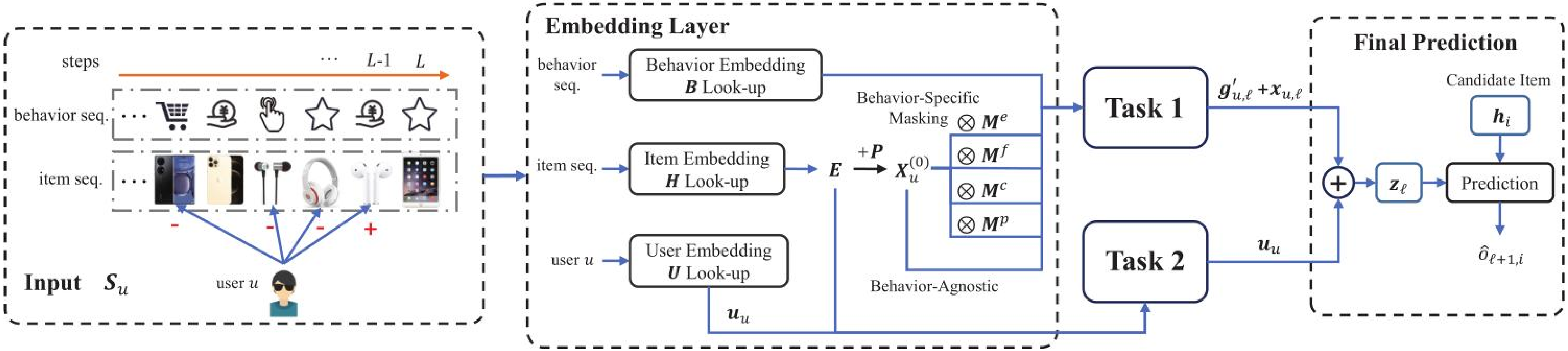
$$N_u = (softmax(\frac{Q''K''^T}{\sqrt{d}}) \otimes A)V'' \tag{13}$$

$$g(\boldsymbol{h}_y) = \exp\left(\boldsymbol{u}_u \boldsymbol{W} \boldsymbol{h}_y^T / \rho\right) \tag{14}$$

$$g(\boldsymbol{n}_{u,t}) = \exp\left(\boldsymbol{u}_u \boldsymbol{W} \boldsymbol{n}_{u,t}^T / \rho\right) \tag{15}$$

$$\mathcal{L}_2 = -\sum_{u \in \mathcal{U}} \sum_{y \in \mathcal{Y}_u} \log \frac{g(\boldsymbol{h}_y)}{g(\boldsymbol{h}_y) + \sum_{t \in \mathcal{T}(u,y)} g(\boldsymbol{n}_{u,t})} \tag{16}$$

Method



$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2 \tag{17}$$

$$\hat{o}_{\ell+1,i} = z_\ell (h_i)^T \tag{18}$$

Experiments

Dataset	# Users	# Items	Avg. Length	Behavior set
UB	20,858	30,853	33.71	$\{e, f, c, p\}$
Tmall	17,209	16,174	48.60	$\{e, f, p\}$

Table 2: Statistics of the processed datasets, where Avg. Length denote the average length of users' interaction sequences in the datasets, and e, f, c and p denote examination, add-to-favorite, add-to-cart and purchase, respectively.

Experiments

Model	UB					Tmall				
	Rec@1	Rec@5	NDCG@5	Rec@10	NDCG@10	Rec@1	Rec@5	NDCG@5	Rec@10	NDCG@10
BPRMF	0.086	0.211	0.149	0.309	0.181	0.050	0.166	0.108	0.266	0.140
FISM	0.095	0.246	0.172	0.362	0.209	0.046	0.171	0.109	0.273	0.142
MB-GMN	0.094	0.251	0.175	0.364	0.208	0.061	0.211	0.136	0.342	0.181
VAE++	0.139	0.290	0.215	0.398	0.250	0.088	0.238	0.167	0.360	0.203
FPMC	0.104	0.257	0.182	0.372	0.219	0.112	0.277	0.196	0.394	0.234
Fossil	0.085	0.219	0.153	0.319	0.185	0.094	0.244	0.169	0.356	0.206
GRU4Rec+	0.225	0.367	0.300	0.451	0.327	0.195	0.381	0.291	0.493	0.327
Caser	0.187	0.353	0.274	0.444	0.304	0.161	0.384	0.275	0.514	0.317
SASRec	0.226	0.446	0.341	0.556	0.376	0.210	0.481	0.351	0.616	0.395
FISSA	0.224	<u>0.493</u>	<u>0.364</u>	<u>0.622</u>	<u>0.405</u>	0.195	0.484	0.344	<u>0.634</u>	0.392
RIB	0.214	0.390	0.306	0.488	0.337	0.205	0.425	0.319	0.547	0.359
BINN	0.223	0.402	0.316	0.505	0.349	<u>0.223</u>	0.434	0.332	0.552	0.370
MGNN-SPred	0.146	0.291	0.220	0.392	0.253	0.165	0.391	0.282	0.521	0.324
M-SR	0.224	0.401	0.316	0.500	0.348	0.217	0.426	0.325	0.547	0.365
ASLI	<u>0.230</u>	0.452	0.347	0.562	0.382	0.215	<u>0.490</u>	<u>0.359</u>	0.623	<u>0.402</u>
NextIP	**0.247**	**0.509**	**0.384**	**0.632**	**0.423**	**0.246**	**0.548**	**0.403**	**0.681**	**0.446**

Table 3: Recommendation performance of our NextIP and four groups of baselines on UB and Tmall. Note that the best one of each column is marked in bold, and the second best result is underlined.

Experiments

Dataset	Metric	BPRMF	VAE++	SASRec	ASLI	NextIP
UB	Rec@5	0.0143	0.0377	<u>0.0436</u>	0.0423	**0.0448**
	NDCG@5	0.0086	0.0250	<u>0.0224</u>	0.0221	**0.0231**
	Rec@10	0.0281	0.0564	<u>0.0766</u>	0.0731	**0.0790**
	NDCG@10	0.0130	0.0310	<u>0.0331</u>	0.0320	**0.0340**
Tmall	Rec@5	0.0094	0.0255	0.0488	<u>0.0514</u>	**0.0542**
	NDCG@5	0.0057	0.0175	0.0271	<u>0.0283</u>	**0.0314**
	Rec@10	0.0189	0.0387	0.0821	<u>0.0859</u>	**0.0896**
	NDCG@10	0.0087	0.0217	0.0379	<u>0.0394</u>	**0.0428**

Table 4: Recommendation performance of our NextIP and four representative baselines on UB and Tmall under the full-ranking setting, in which all items are considered as candidates. Note that the best one of each column is marked in bold, and the second best result is underlined.

Experiments

Architecture	UB	Tmall
NextIP(w/o TBCG&BSA)	0.556	0.616
NextIP(w/o TBCG)	0.577	0.635
NextIP(w/o BSA)	0.624	0.678
NextIP(w/o BSA & $\boldsymbol{g}_{u,\ell}$ in TBCG)	0.557	0.634
NextIP(w/o BSA & $\boldsymbol{x}^*_{u,\ell}$ in TBCG)	0.570	0.648
NextIP	**0.632**	**0.681**

Table 5: Recommendation performance (Rec@10) of our NextIP with different architectures on UB and Tmall for ablation studies.

Experiments

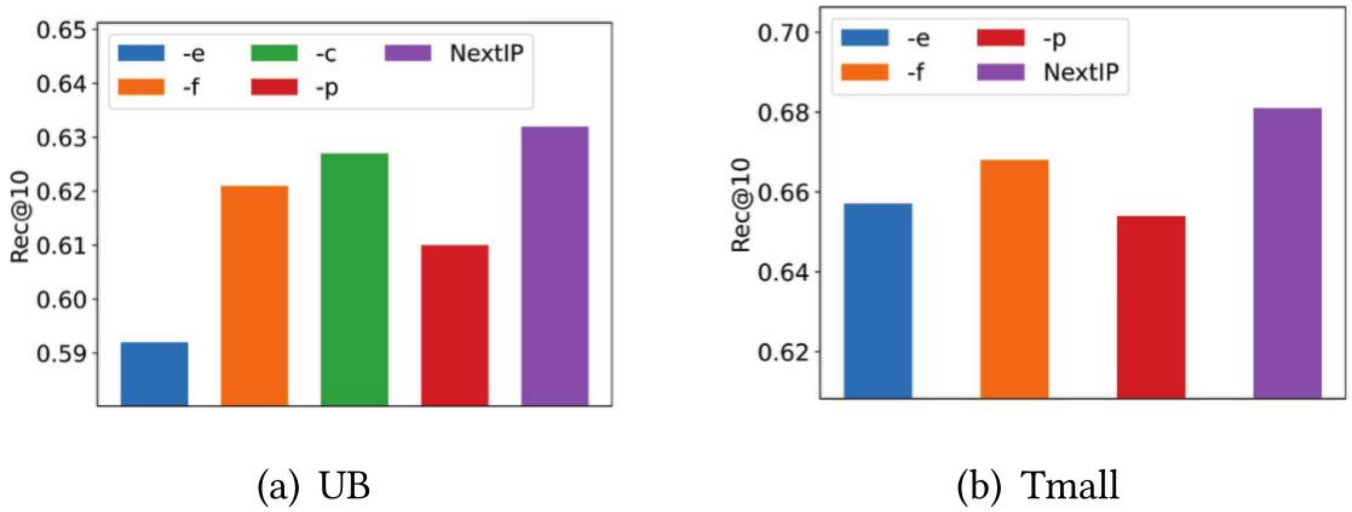
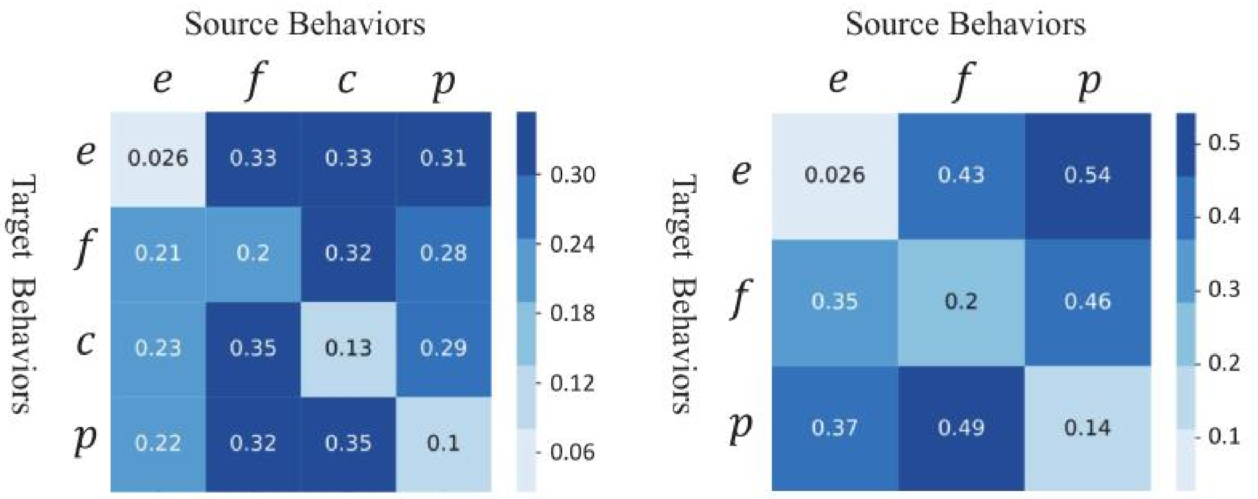


Figure 3: Recommendation performance (Rec@10) of our NextIP and its variants by removing different behavior-specific item sub-sequences on Tmall and UB.

Experiments



(a) **UB** (b) **Tmall**

Figure 4: Visualization of the attention scores of different source behaviors to different target behaviors in target-behavior aware context aggregator (TBCG) of our NextIP on UB and Tmall.

Thanks