Towards Universal Sequence Representation Learning for Recommender Systems

Advisor: Jia-Ling, Koh
Speaker: Hsiang-Hui Chou
Source: KDD’22
Date:2022/11/07

Outline

- Introduction
- Method
- Experiment
- Conclusion

Sequential recommendation

Input
user interaction sequence $s = \{i_1, i_2, \ldots i_n\}$

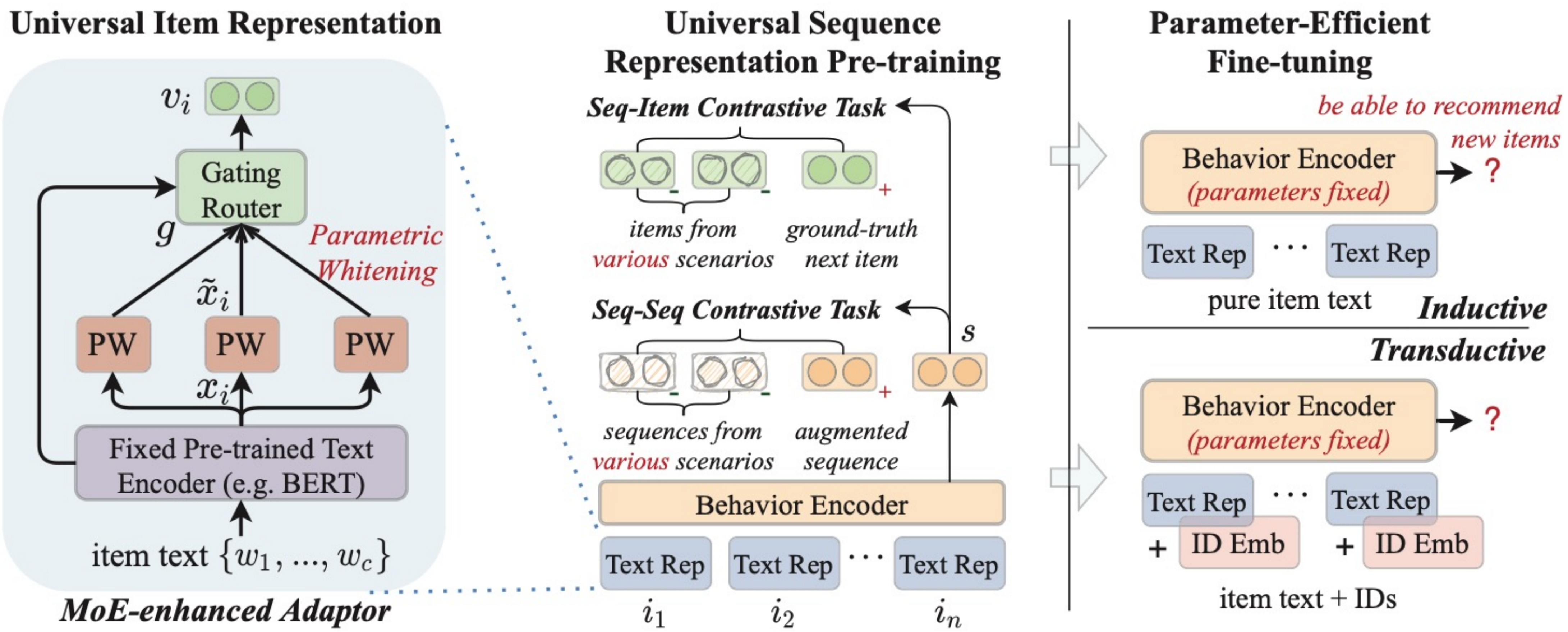
each interacted item i associated with a unique item ID and
description text $t_i = \{w_1, w_2, \ldots w_c\}$

Output
probability

Outline

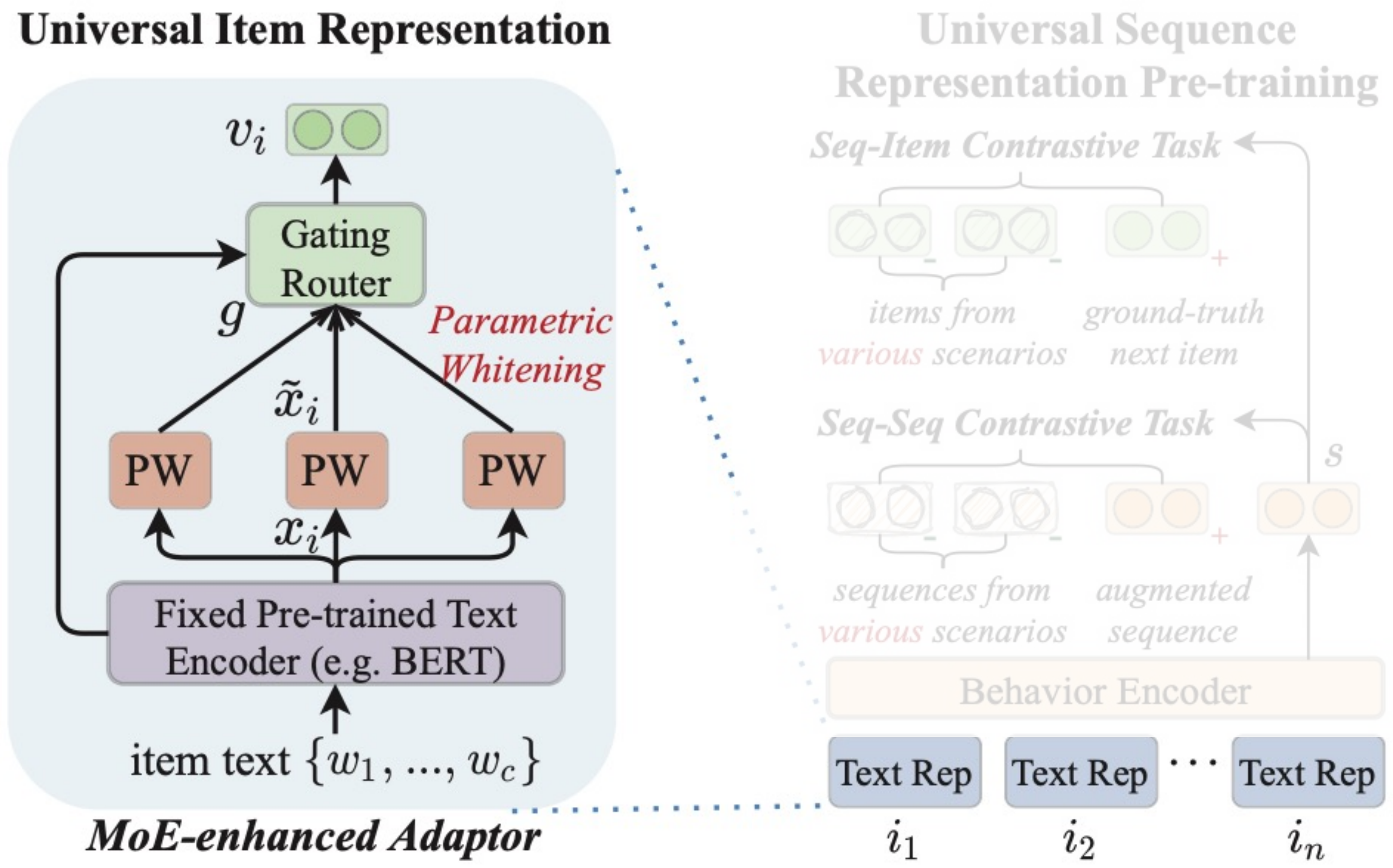
- Introduction
- Method
- Experiment
- Conclusion

Universal sequence representation learning approach (UniSRec)



Universal Textual Item Representation

-Semantic Transformation via Parametric Whitening



$$\boldsymbol{x}_i = \mathrm{BERT}([[\mathtt{CLS}]; w_1, \ldots, w_c])$$

$$\widetilde{\boldsymbol{x}}_i = (\boldsymbol{x}_i - \boldsymbol{b}) \cdot \boldsymbol{W}_1$$

Bert whitening

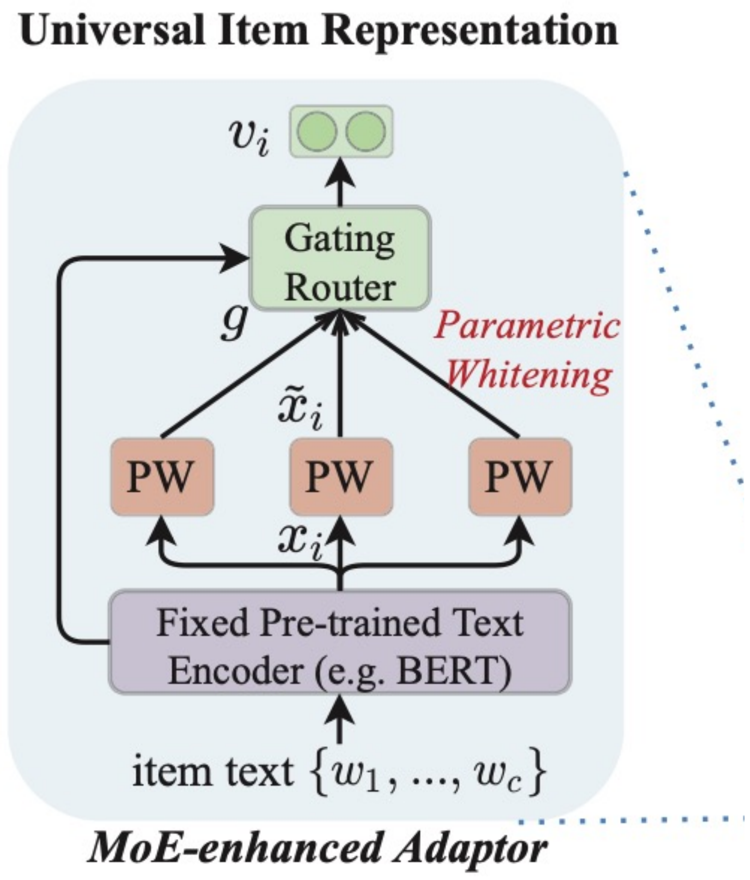
Algorithm 1 Whitening-k Workflow

Input: Existing embeddings $\{x_i\}_{i=1}^N$ and reserved dimensionality k

1: compute μ and Σ of $\{x_i\}_{i=1}^N$
2: compute $U, \Lambda, U^T = \text{SVD}(\Sigma)$
3: compute $W = (U\sqrt{\Lambda^{-1}})[:, :k]$
4: **for** $i = 1, 2, \cdots, N$ **do**
5: $\quad \widetilde{x}_i = (x_i - \mu)W$
6: **end for**

Output: Transformed embeddings $\{\widetilde{x}_i\}_{i=1}^N$

MoE-enhanced Adaptor



$$\boldsymbol{x}_i = \mathrm{BERT}([[\mathtt{CLS}]; w_1, \ldots, w_c])$$

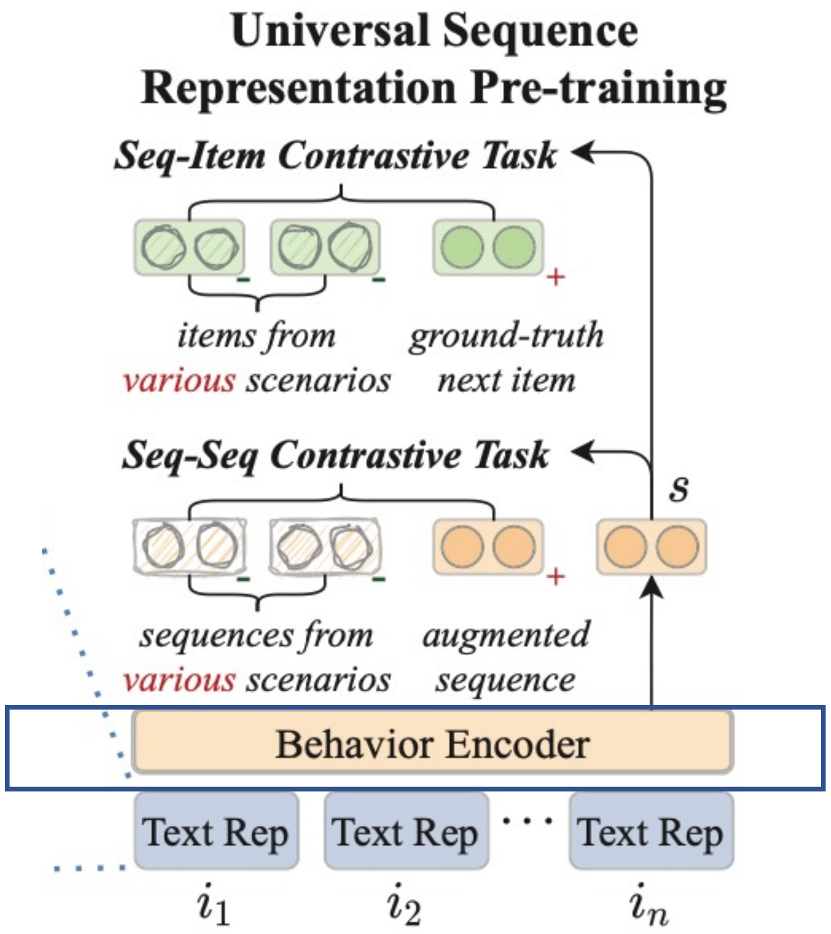
$$\widetilde{\boldsymbol{x}}_i = (\boldsymbol{x}_i - \boldsymbol{b}) \cdot \boldsymbol{W}_1$$

$$\boldsymbol{v}_i = \sum_{k=1}^{G} g_k \cdot \widetilde{\boldsymbol{x}}_i^{(k)}$$

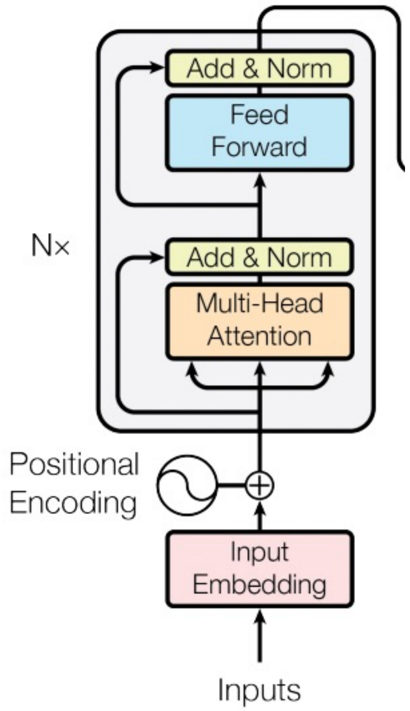
$$\boldsymbol{g} = \mathrm{Softmax}\left(\boldsymbol{x}_i \cdot \boldsymbol{W}_2 + \boldsymbol{\delta}\right),$$
$$\boldsymbol{\delta} = \mathrm{Norm}() \cdot \mathrm{Softplus}\left(\boldsymbol{x}_i \cdot \boldsymbol{W}_3\right).$$

Universal Sequence Representation

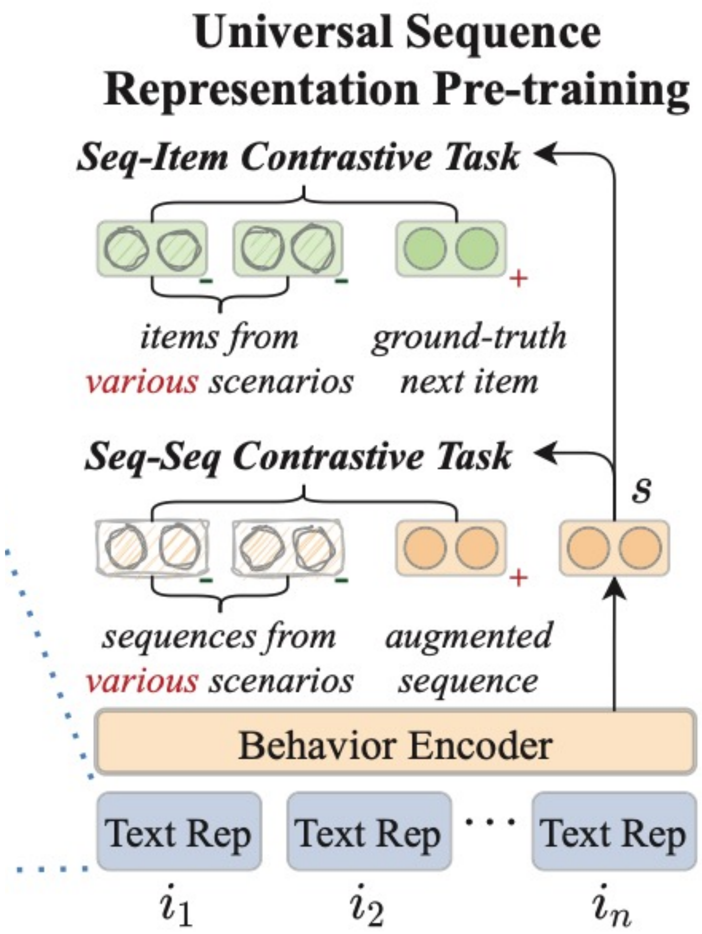
-Self-attentive Sequence Encoding



$$f_j^0 = v_i + p_j,$$
$$F^{l+1} = \mathrm{FFN}(\mathrm{MHAttn}(F^l)),$$



Universal Sequence Representation



Sequence-item contrastive task

$$\ell_{S-I} = -\sum_{j=1}^{B} \log \frac{\exp\left(\boldsymbol{s}_j \cdot \boldsymbol{v}_j / \tau\right)}{\sum_{j'=1}^{B} \exp\left(\boldsymbol{s}_j \cdot \boldsymbol{v}_{j'} / \tau\right)},$$

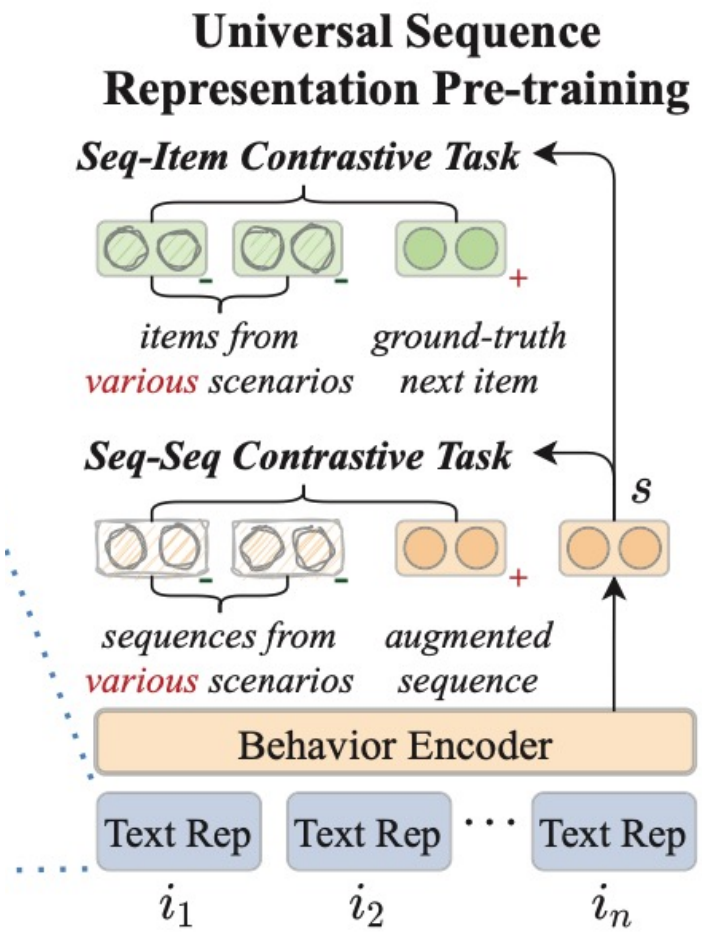
Sequence-sequence contrastive task

$$\ell_{S-S} = -\sum_{j=1}^{B} \log \frac{\exp\left(\boldsymbol{s}_j \cdot \widetilde{\boldsymbol{s}}_j / \tau\right)}{\sum_{j'=1}^{B} \exp\left(\boldsymbol{s}_j \cdot \boldsymbol{s}_{j'} / \tau\right)}.$$

Universal Sequence Representation

$$\ell_{S-I} = -\sum_{j=1}^{B} \log \frac{\exp\left(\boldsymbol{s}_j \cdot \boldsymbol{v}_j / \tau\right)}{\sum_{j'=1}^{B} \exp\left(\boldsymbol{s}_j \cdot \boldsymbol{v}_{j'} / \tau\right)},$$

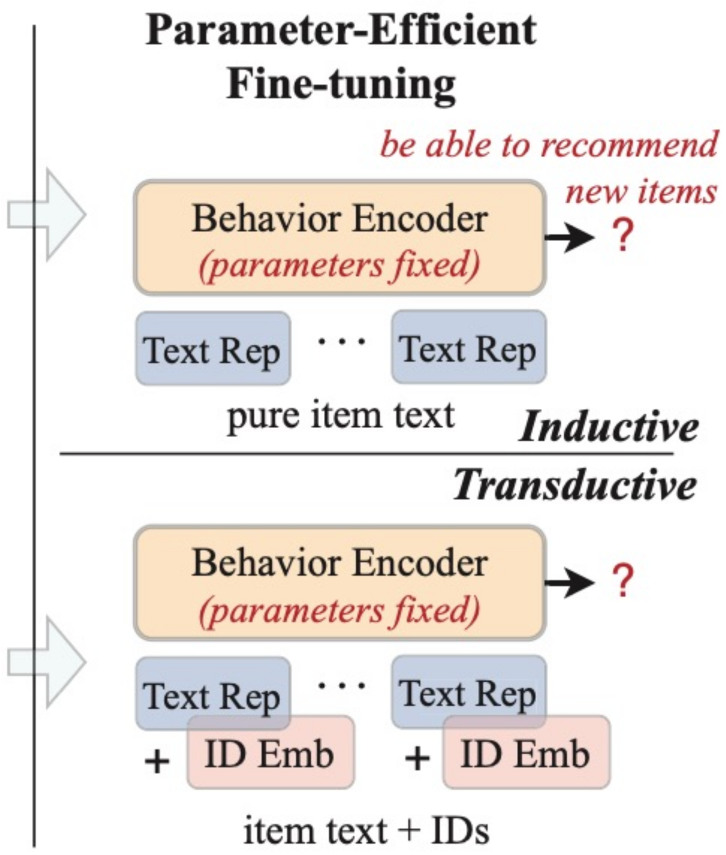
$$\ell_{S-S} = -\sum_{j=1}^{B} \log \frac{\exp\left(\boldsymbol{s}_j \cdot \widetilde{\boldsymbol{s}}_j / \tau\right)}{\sum_{j'=1}^{B} \exp\left(\boldsymbol{s}_j \cdot \boldsymbol{s}_{j'} / \tau\right)}.$$



Multi-task learning

$$\mathcal{L}_{\mathrm{PT}} = \ell_{S-I} + \lambda \cdot \ell_{S-S},$$

Parameter-Efficient Fine-tuning



Inductive setting

$$P_I(i_{t+1}|s) = \text{Softmax}(\boldsymbol{s} \cdot \boldsymbol{v}_{i_{t+1}}),$$

Transductive setting

$$P_T(i_{t+1}|s) = \text{Softmax}\left(\widetilde{\boldsymbol{s}} \cdot (\boldsymbol{v}_{i_{t+1}} + \boldsymbol{e}_{i_{t+1}})\right)$$

Outline

- Introduction
- Method
- Experiment
- Conclusion

Dataset

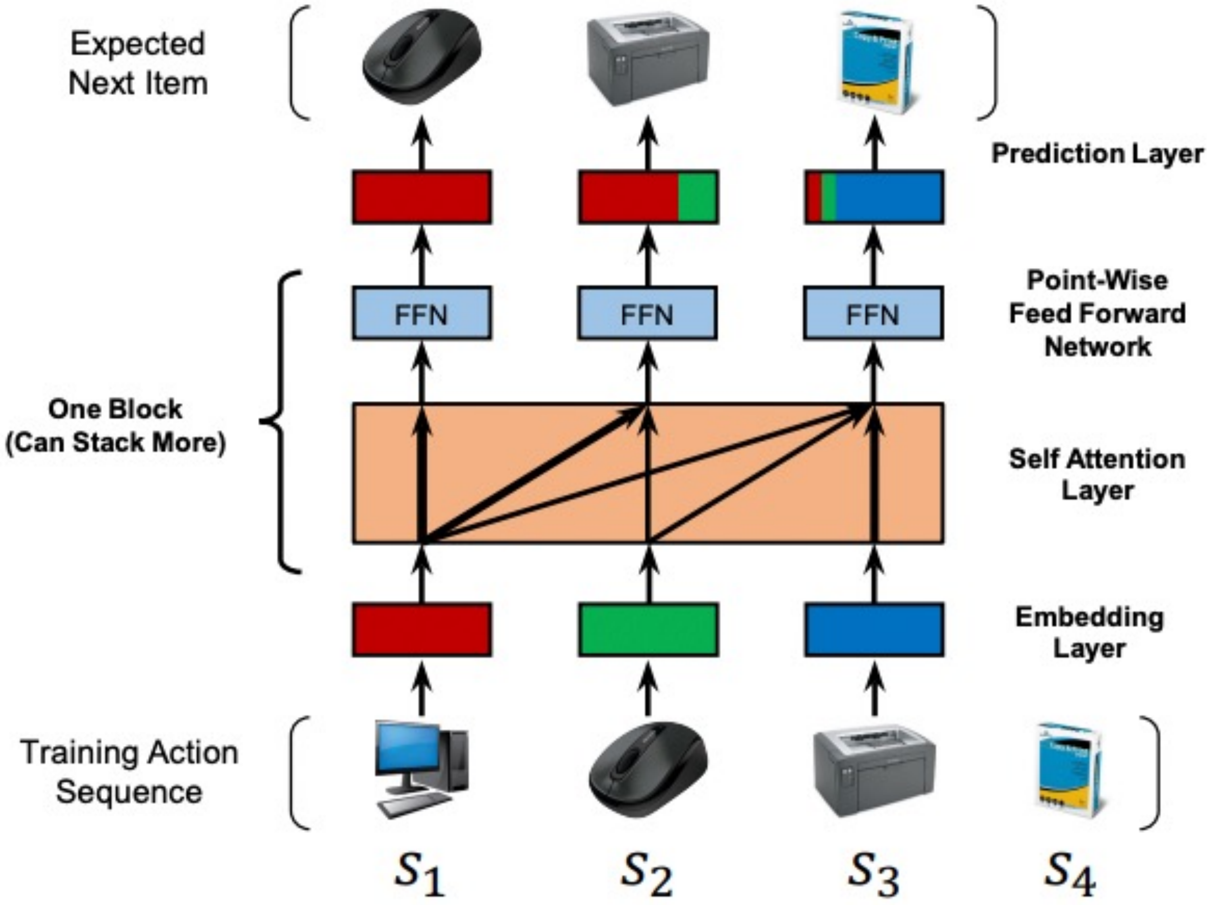
Datasets	#Users	#Items	#Inters.	Avg. n	Avg. c
Pre-trained	1,361,408	446,975	14,029,229	13.51	139.34
- Food	115,349	39,670	1,027,413	8.91	153.40
- CDs	94,010	64,439	1,118,563	12.64	80.43
- Kindle	138,436	98,111	2,204,596	15.93	141.70
- Movies	281,700	59.203	3,226,731	11.45	97.54
- Home	731,913	185,552	6,451,926	8.82	168.89
Scientific	8,442	4,385	59,427	7.04	182.87
Pantry	13,101	4,898	126,962	9.69	83.17
Instruments	24,962	9,964	208,926	8.37	165.18
Arts	45,486	21,019	395,150	8.69	155.57
Office	87,436	25,986	684,837	7.84	193.22
Online Retail	16,520	3,469	519,906	26.90	27.80

Cross-domain

Cross-platform

Scenario	Dataset	Metric	SASRec	BERT4Rec	FDSA	S^3-Rec	CCDR	RecGURU	ZESRec	UniSRec$_t$	UniSRec$_{t+ID}$	Improv.
Cross-Domain	Scientific	Recall@10	0.1080	0.0488	0.0899	0.0525	0.0695	0.1023	0.0851	0.1188*	**0.1235***	+14.35%
		NDCG@10	0.0553	0.0243	0.0580	0.0275	0.0340	0.0572	0.0475	**0.0641***	0.0634*	+10.52%
		Recall@50	0.2042	0.1185	0.1732	0.1418	0.1647	0.2022	0.1746	0.2394*	**0.2473***	+21.11%
		NDCG@50	0.0760	0.0393	0.0759	0.0468	0.0546	0.0786	0.0670	0.0903*	**0.0904***	+15.01%
	Pantry	Recall@10	0.0501	0.0308	0.0395	0.0444	0.0408	0.0469	0.0454	0.0636*	**0.0693***	+38.32%
		NDCG@10	0.0218	0.0152	0.0209	0.0214	0.0203	0.0209	0.0230	0.0306*	**0.0311***	+35.22%
		Recall@50	0.1322	0.1030	0.1151	0.1315	0.1262	0.1269	0.1141	0.1658*	**0.1827***	+38.20%
		NDCG@50	0.0394	0.0305	0.0370	0.0400	0.0385	0.0379	0.0378	0.0527*	**0.0556***	+39.00%
	Instruments	Recall@10	0.1118	0.0813	0.1070	0.1056	0.0848	0.1113	0.0783	0.1189*	**0.1267***	+13.33%
		NDCG@10	0.0612	0.0620	**0.0796**	0.0713	0.0451	0.0681	0.0497	0.0680	0.0748*	–
		Recall@50	0.2106	0.1454	0.1890	0.1927	0.1753	0.2068	0.1387	0.2255*	**0.2387***	+13.34%
		NDCG@50	0.0826	0.0756	0.0972	0.0901	0.0647	0.0887	0.0627	0.0912	**0.0991***	+1.95%
	Arts	Recall@10	0.1108	0.0722	0.1002	0.1003	0.0671	0.1084	0.0664	0.1066	**0.1239***	+11.82%
		NDCG@10	0.0587	0.0479	**0.0714**	0.0601	0.0348	0.0651	0.0375	0.0586	0.0712	–
		Recall@50	0.2030	0.1367	0.1779	0.1888	0.1478	0.1979	0.1323	0.2049*	**0.2347***	+15.62%
		NDCG@50	0.0788	0.0619	0.0883	0.0793	0.0523	0.0845	0.0518	0.0799	**0.0955***	+8.15%
	Office	Recall@10	0.1056	0.0825	0.1118	0.1030	0.0549	0.1145	0.0641	0.1013	**0.1280***	+11.79%
		NDCG@10	0.0710	0.0634	**0.0868**	0.0653	0.0290	0.0768	0.0391	0.0619	0.0831	–
		Recall@50	0.1627	0.1227	0.1665	0.1613	0.1095	0.1757	0.1113	0.1702	**0.2016***	+14.74%
		NDCG@50	0.0835	0.0721	0.0987	0.0780	0.0409	0.0901	0.0493	0.0769	**0.0991**	+0.41%
Cross-Platform	Online Retail	Recall@10	0.1460	0.1349	0.1490	0.1418	0.1347	0.1467	0.1103	0.1449	**0.1537***	+3.15%
		NDCG@10	0.0675	0.0653	0.0719	0.0654	0.0620	0.0658	0.0535	0.0677	**0.0724**	+0.70%
		Recall@50	0.3872	0.3540	0.3802	0.3702	0.3587	**0.3885**	0.2750	0.3604	**0.3885**	0.00%
		NDCG@50	0.1201	0.1131	0.1223	0.1154	0.1108	0.1188	0.0896	0.1149	**0.1239***	+1.31%

SASRec



Expected Next Item
Prediction Layer
Point-Wise Feed Forward Network
FFN
FFN
FFN
One Block (Can Stack More)
Self Attention Layer
Embedding Layer
Training Action Sequence
s_1
s_2
s_3
s_4

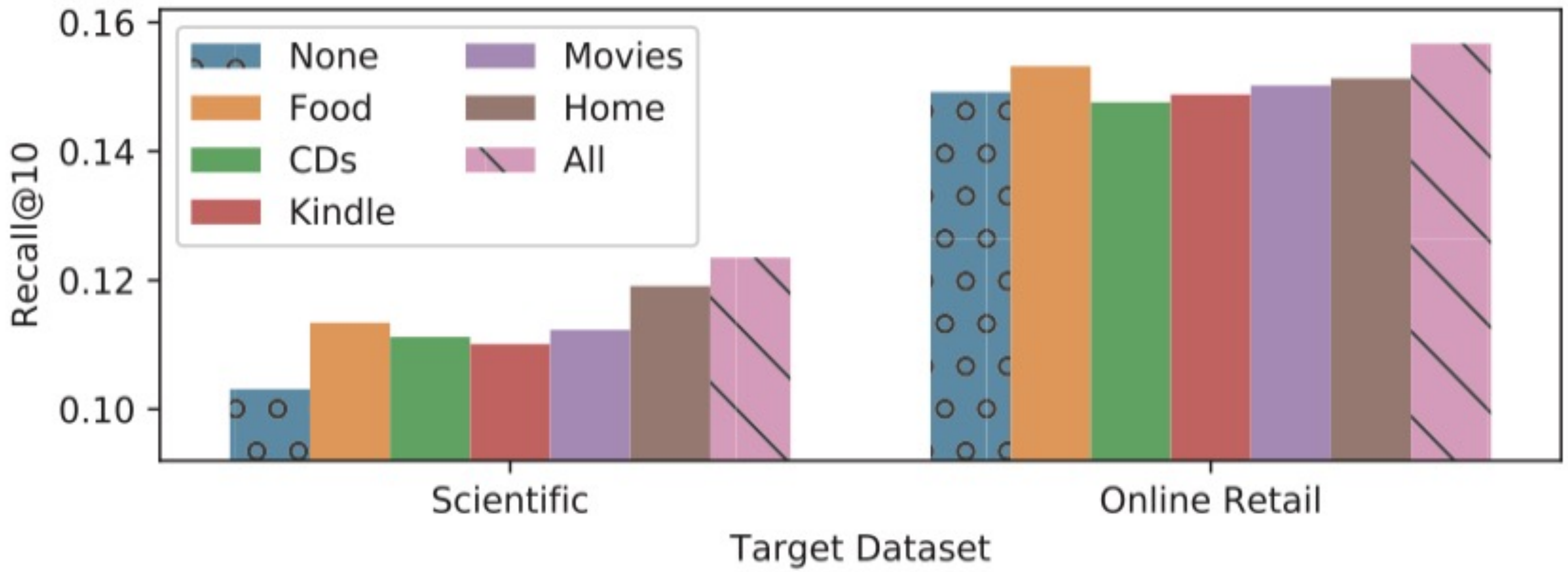


Figure 2: Performance comparison w.r.t. different pre-training datasets on "Scientific" and "Online Retail". "All" denotes the model pre-trained on all five datasets, and "None" denotes the training from scratch.

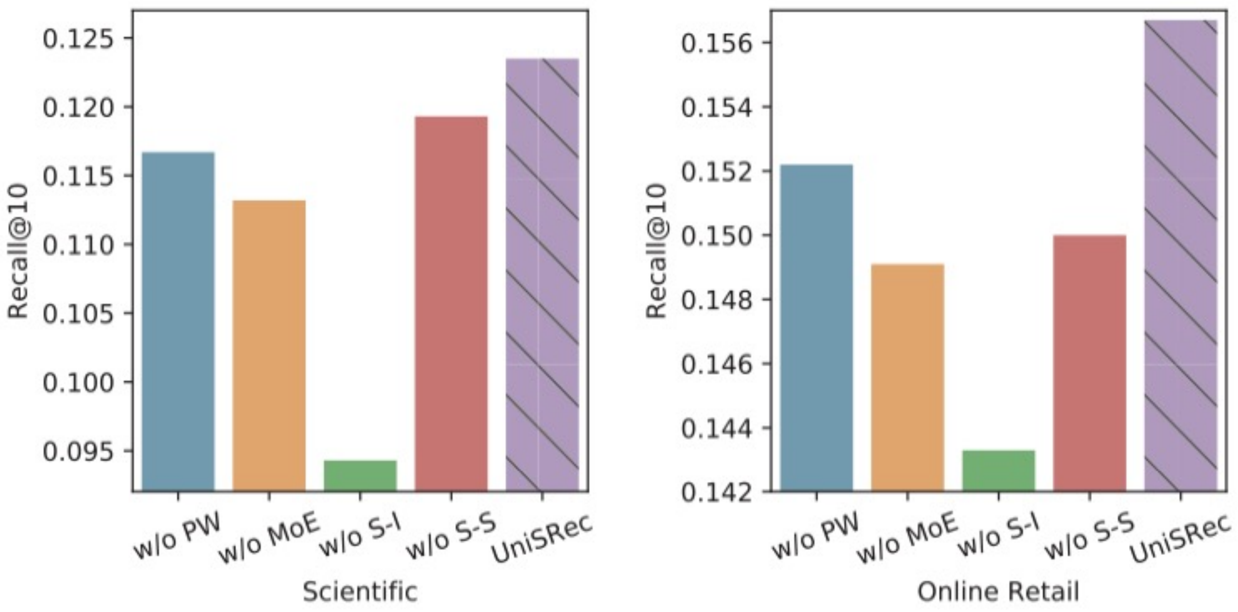
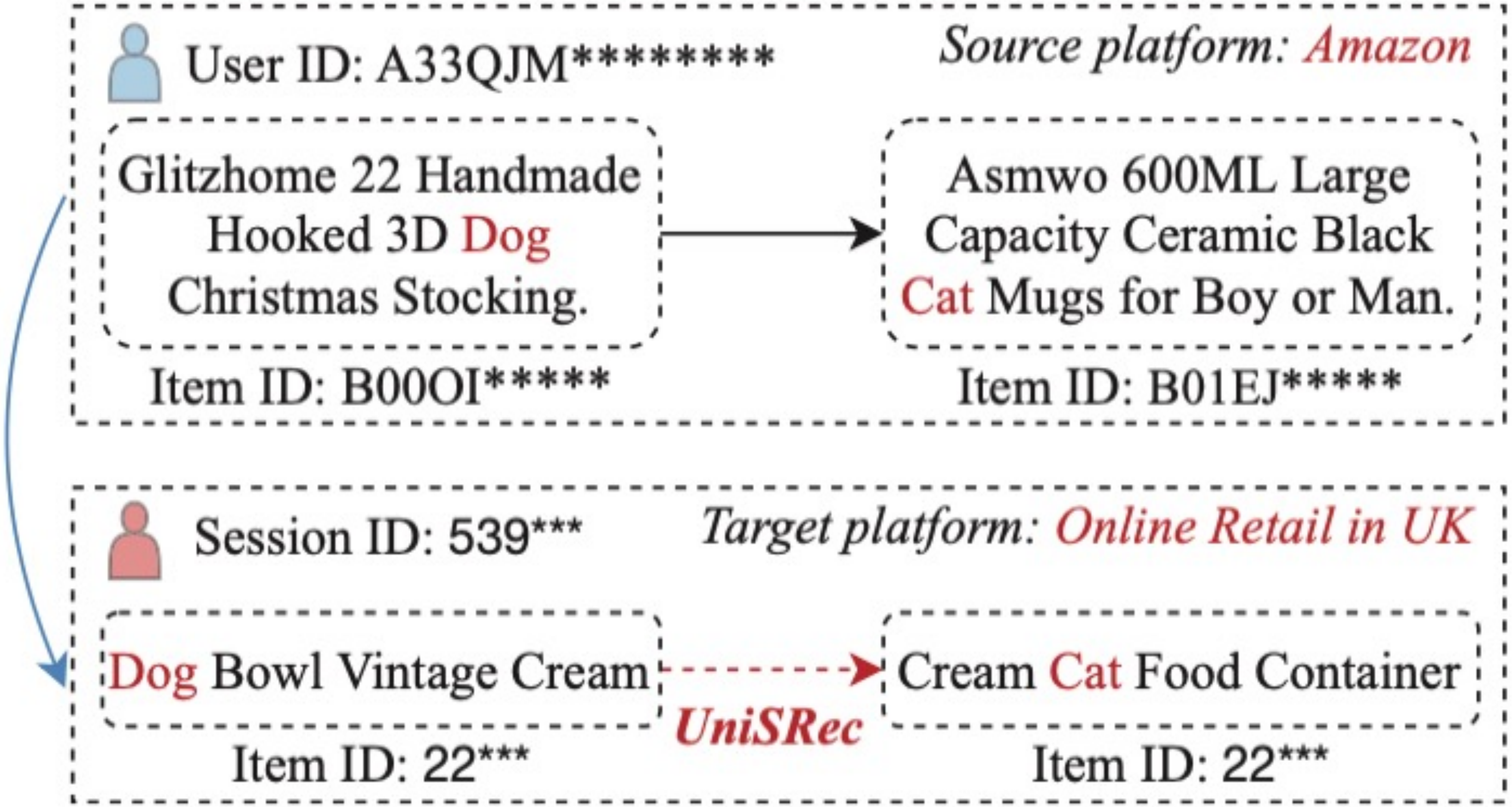


Figure 3: Ablation study of UniSRec variants on "Scientific" and "Online Retail".


User ID: A33QJM********
Source platform: Amazon
Glitzhome 22 Handmade Hooked 3D Dog Christmas Stocking.
Item ID: B00OI*****
Asmwo 600ML Large Capacity Ceramic Black Cat Mugs for Boy or Man.
Item ID: B01EJ*****
Session ID: 539***
Target platform: Online Retail in UK
Dog Bowl Vintage Cream
Item ID: 22***
UniSRec
Cream Cat Food Container
Item ID: 22***

Outline

- Introduction
- Method
- Experiment
- Conclusion

Conclusion

- UniSRec utilizes item texts to learn more transferable representations for sequential recommendation
- design a lightweight architecture based on parametric whitening and MoE-enhanced adaptor to learn the **universal item representations**
- design two contrastive pre-training tasks to learn **universal sequence representations** from multi-domain sequences