Towards Universal Sequence Representation Learning for Recommender Systems

Yupeng Hou[*†]
houyupeng@ruc.edu.cn
Gaoling School of Artificial Intelligence, Renmin University of China

Shanlei Mu[*]
slmu@ruc.edu.cn
School of Information, Renmin University of China

Wayne Xin Zhao[†‡✉]
batmanfly@gmail.com
Gaoling School of Artificial Intelligence, Renmin University of China

Yaliang Li
yaliang.li@alibaba-inc.com
Alibaba Group

Bolin Ding
bolin.ding@alibaba-inc.com
Alibaba Group

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

code:https://github.com/RUCAIBox/UniSRec.

KDD 2022

Reported by Minqin Li

Introduction

First,the textual semantic space is not directly suited for the recommendation tasks. It is not clear how to model and utilize item texts for improving the recommendation performance, since directly introducing raw textual representations as additional features may lead to suboptimal results.

Second, it is difficult to leverage multi-domain data for improving the target domain, where the seesaw phenomenon (referring to learning from multiple kinds of domain-specific patterns is conflict or oscillating) often appears.

To address the above issues, we propose the universal sequence representation learning approach, named UniSRec.

Method

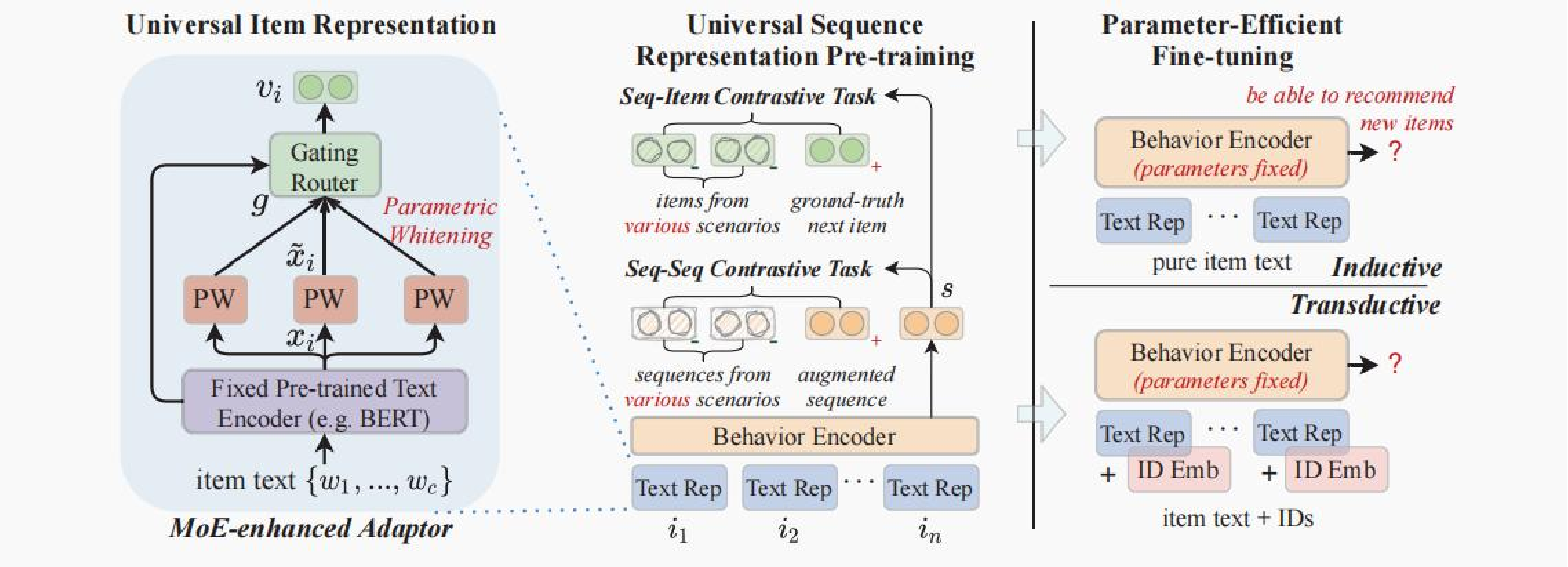

Figure 1: The overall framework of the proposed universal sequence representation learning approach (UniSRec).

Method

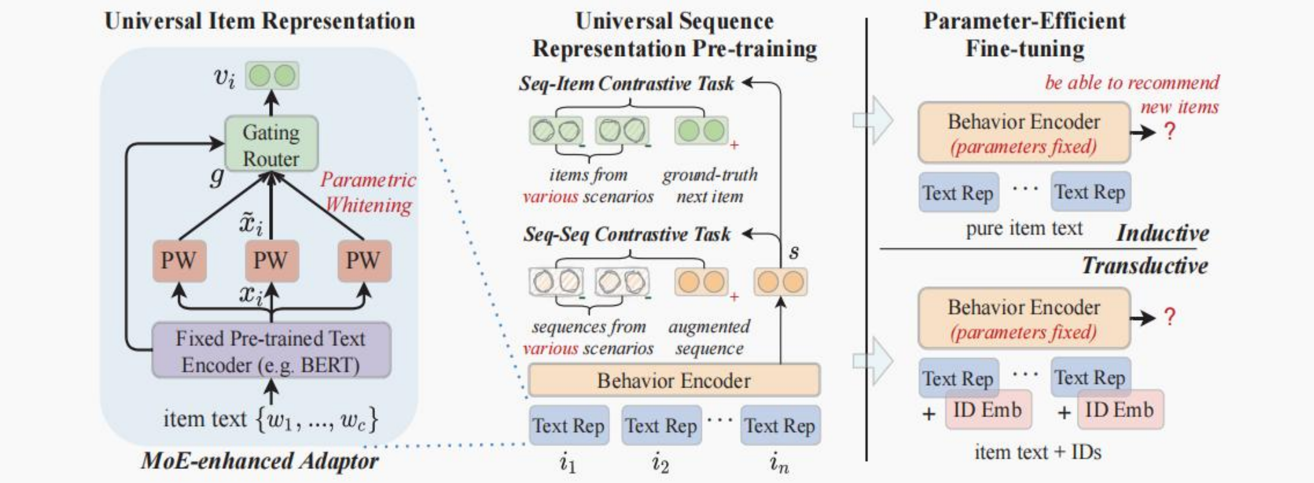

Figure 1: The overall framework of the proposed universal sequence representation learning approach (UniSRec).

$$x_i = \mathrm{BERT}([[\mathtt{CLS}]; w_1, \ldots, w_c]) \quad (1)$$

$$\tilde{x}_i = (x_i - b) \cdot W_1 \quad (2)$$

$$v_i = \sum_{k=1}^{G} g_k \cdot \tilde{x}_i^{(k)} \quad (3)$$

$$g = \mathrm{Softmax}(x_i \cdot W_2 + \delta) \quad (4)$$

$$\delta = \mathrm{Norm}() \cdot \mathrm{Softplus}(x_i \cdot W_3) \quad (5)$$

$$f_j^0 = v_i + p_j \quad (6)$$

$$F^{l+1} = \mathrm{FFN}(\mathrm{MHAttn}(F^l)) \quad (7)$$

$$F^l = [f_0^l; \ldots; f_n^l]$$

Method

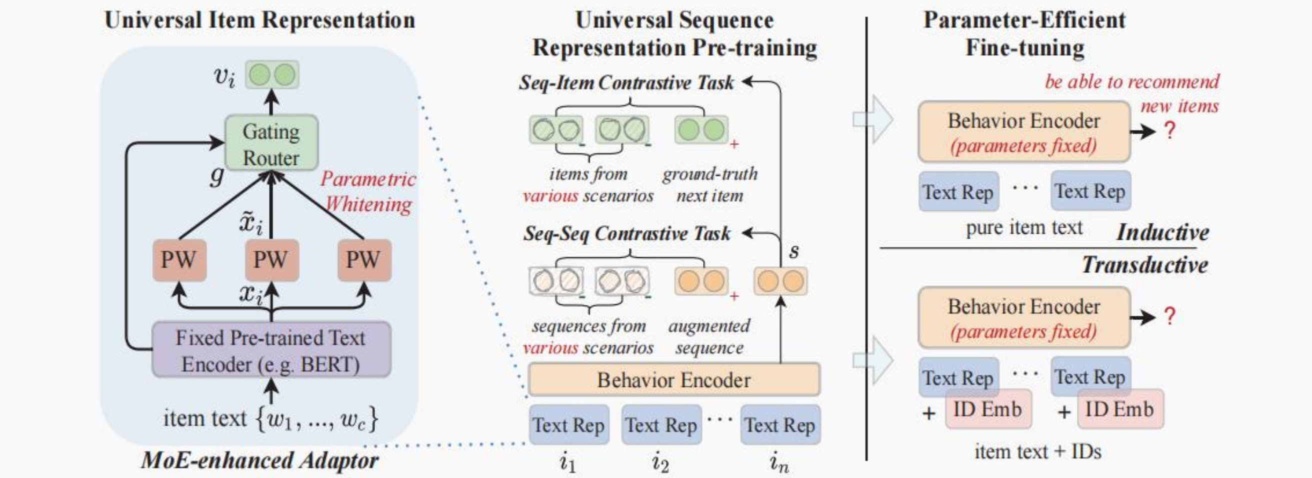

Figure 1: The overall framework of the proposed universal sequence representation learning approach (UniSRec).

$$\mathcal{L}_{\mathrm{PT}} = \ell_{S-I} + \lambda \cdot \ell_{S-S} \quad (10)$$

$$P_I(i_{t+1}|s) = \mathrm{Softmax}(\boldsymbol{s} \cdot \boldsymbol{v}_{i_{t+1}}) \quad (11)$$

$$P_T(i_{t+1}|s) = \mathrm{Softmax}\left(\widetilde{\boldsymbol{s}} \cdot (\boldsymbol{v}_{i_{t+1}} + \boldsymbol{e}_{i_{t+1}})\right) \quad (12)$$

$$\ell_{S-I} = -\sum_{j=1}^{B} \log \frac{\exp(\boldsymbol{s}_j \cdot \boldsymbol{v}_j / \tau)}{\sum_{j'=1}^{B} \exp(\boldsymbol{s}_j \cdot \boldsymbol{v}_{j'} / \tau)} \quad (8)$$

$$\ell_{S-S} = -\sum_{j=1}^{B} \log \frac{\exp(\boldsymbol{s}_j \cdot \widetilde{\boldsymbol{s}}_j / \tau)}{\sum_{j'=1}^{B} \exp(\boldsymbol{s}_j \cdot \boldsymbol{s}_{j'} / \tau)} \quad (9)$$

Experiments

Table 1: Comparison of the transfer learning scenarios and application settings of several approaches. $1 \rightarrow 1$ denotes 1 source domain to 1 target domain, and $M \rightarrow N$ denotes M source domains to N target domains. "Non-OL" denotes that the approach doesn't require overlapped users or items.

Methods	Transfer Learning Scenarios			Application Settings	
	$1 \rightarrow 1$	$M \rightarrow N$	Non-OL	Transductive	Inductive
S^3-Rec [36]	✗	✗	✗	✔	✗
PeterRec [31]	✔	✗	✗	✔	✗
RecGURU [12]	✔	✗	✔	✔	✗
ZESRec [4]	✔	✗	✔	✗	✔
UniSRec (ours)	✔	✔	✔	✔	✔

Experiments

Table 2: Statistics of the datasets after preprocessing. "Avg. n" denotes the average length of item sequences. "Avg. c" denotes the average number of tokens in item text.

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

Experiments

Table 3: Performance comparison of different recommendation models. The best and the second-best performances are denoted in bold and underlined fonts, respectively. "Improv." indicates the relative improvement ratios of the proposed approach over the best performance baselines. "*" denotes that the improvements are significant at the level of 0.01 with paired t-test.

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%

Experiments

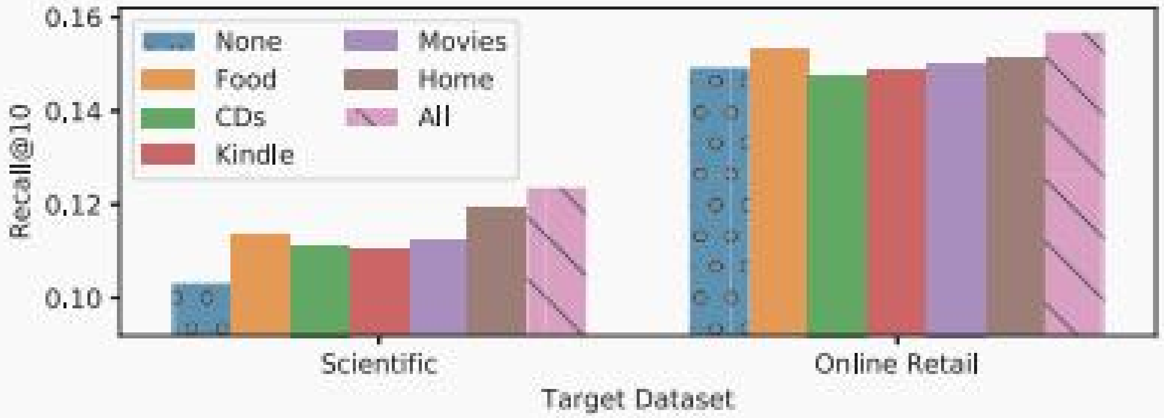

Figure 2: Performance comparison w.r.t. different pre-training datasets on "Scientific" and "Online Retail". "All" denotes the model pre-trained on all five datasets, and "None" denotes the training from scratch.

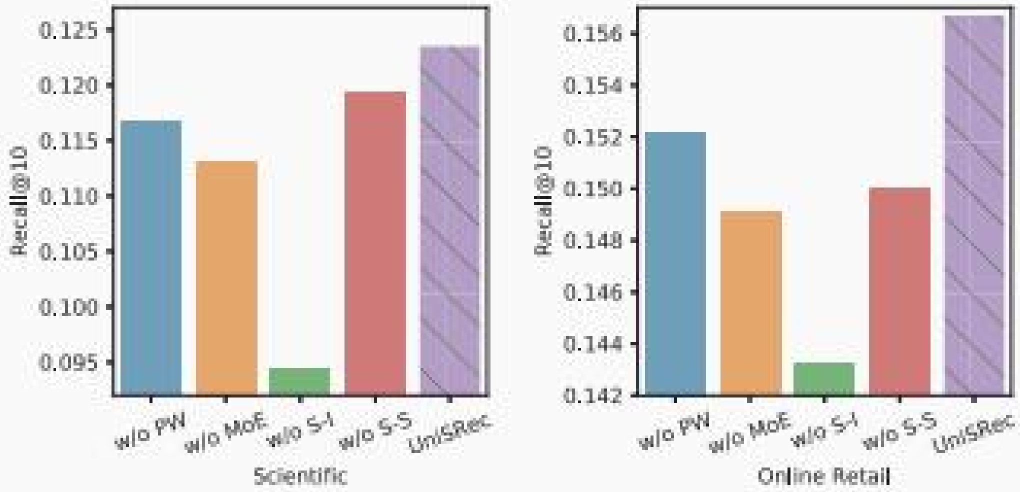

Figure 3: Ablation study of UniSRec variants on "Scientific" and "Online Retail".

Experiments

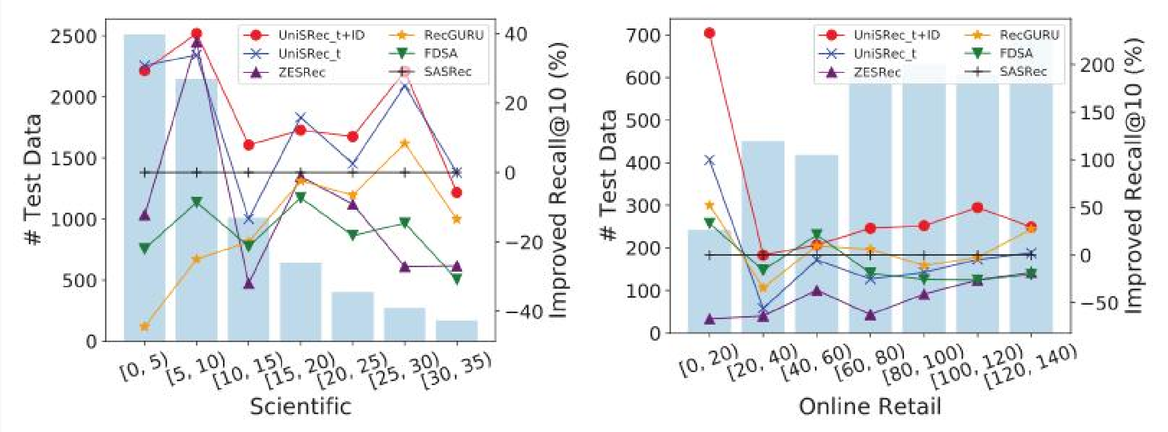

Figure 4: Performance comparison w.r.t. long-tail items on the "Scientific" and "Online Retail" datasets. The bar graph represents the number of interactions in test data for each group. The line chart represents the improvement ratios for Recall@10 compared with SASRec.

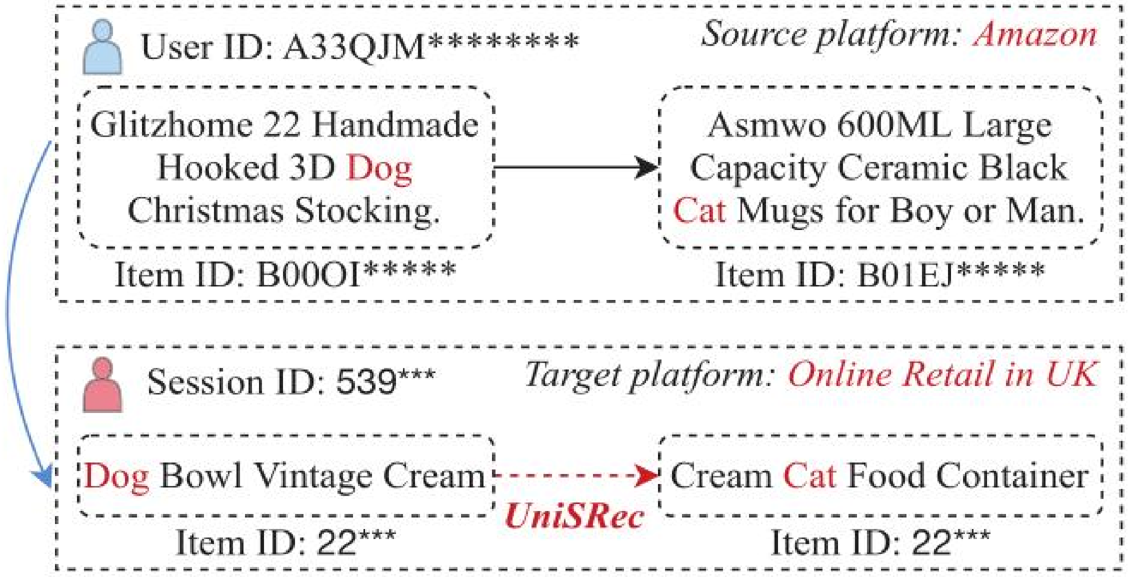

Figure 5: The purchase history of a user in the source platform (top) and the purchase history within an anonymous session in the target platform (bottom). There are naturally no overlapping users and items between the two platforms. This case shows that UniSRec can capture the universal semantic sequence pattern (*e.g.*, "Dog → Cat") from large-scale pre-training, which helps improve the recommendation performance of the target platform.

Thanks