KDD 2022

Towards Universal Sequence Representation Learning for Recommender Systems

Yupeng Hou[1], Shanlei Mu[1], Wayne Xin Zhao [1], Yaliang Li[2], Bolin Ding[2], Ji-Rong Wen[1]

1 Renmin University of China

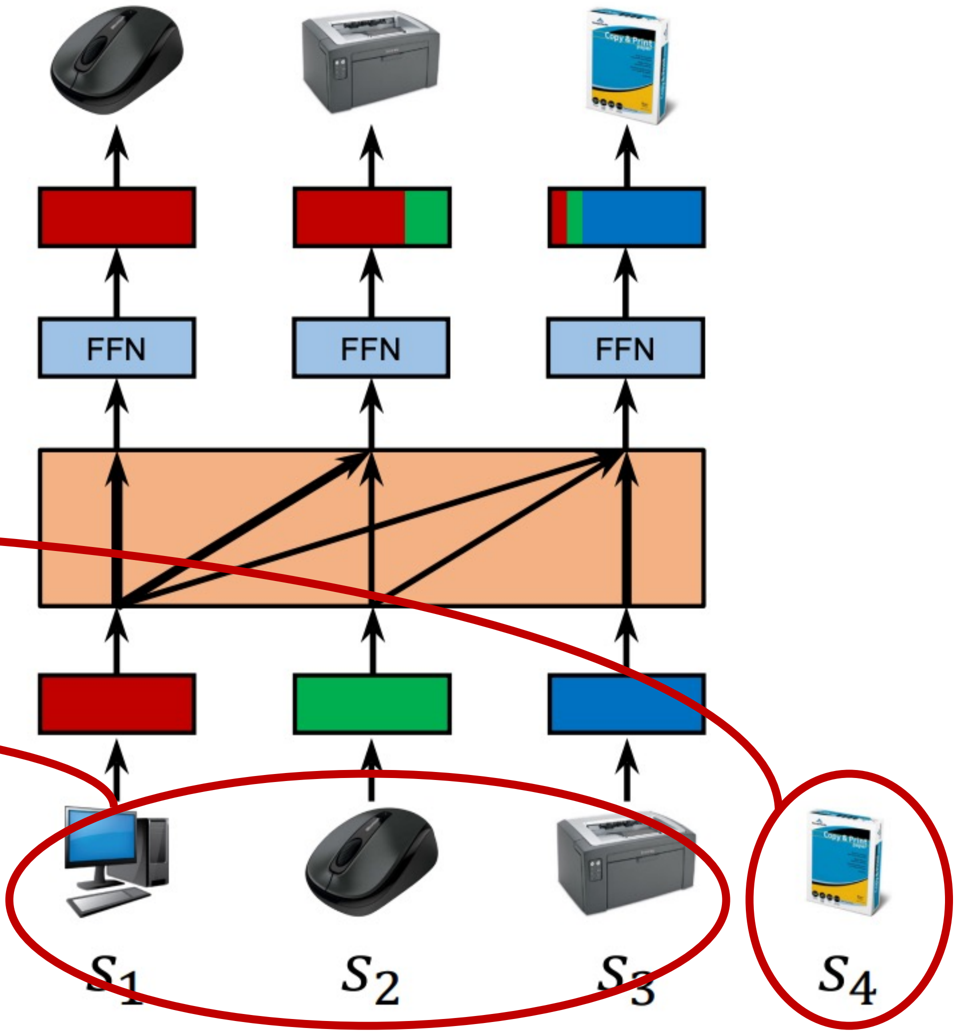
2 Alibaba Group

Motivation

- Sequential Recommendation

Next Item Prediction

Historical Behavior Sequences



Kang and McAuley. Self-Attentive Sequential Recommendation. ICDM 2018.

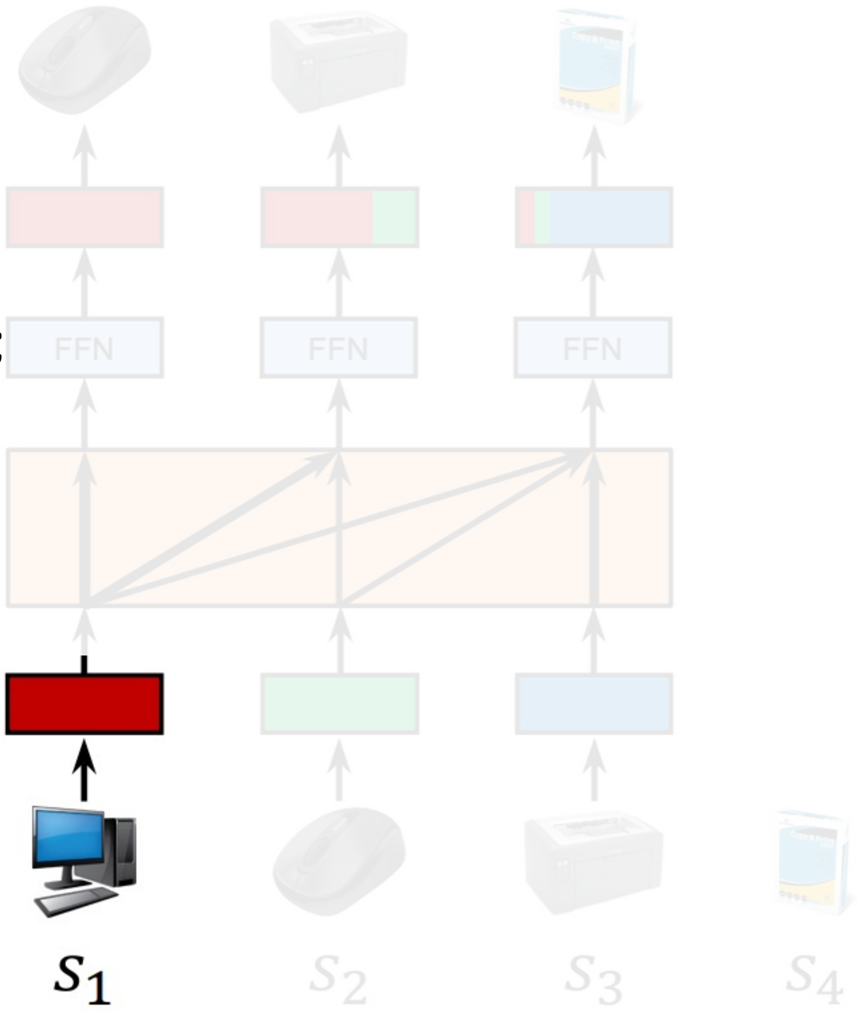
Motivation

- Sequential Recommendation
  - Existing methods: explicit item ID modeling;

Sequence Model

ID Embedding

Item ID, B00TNNWMXI

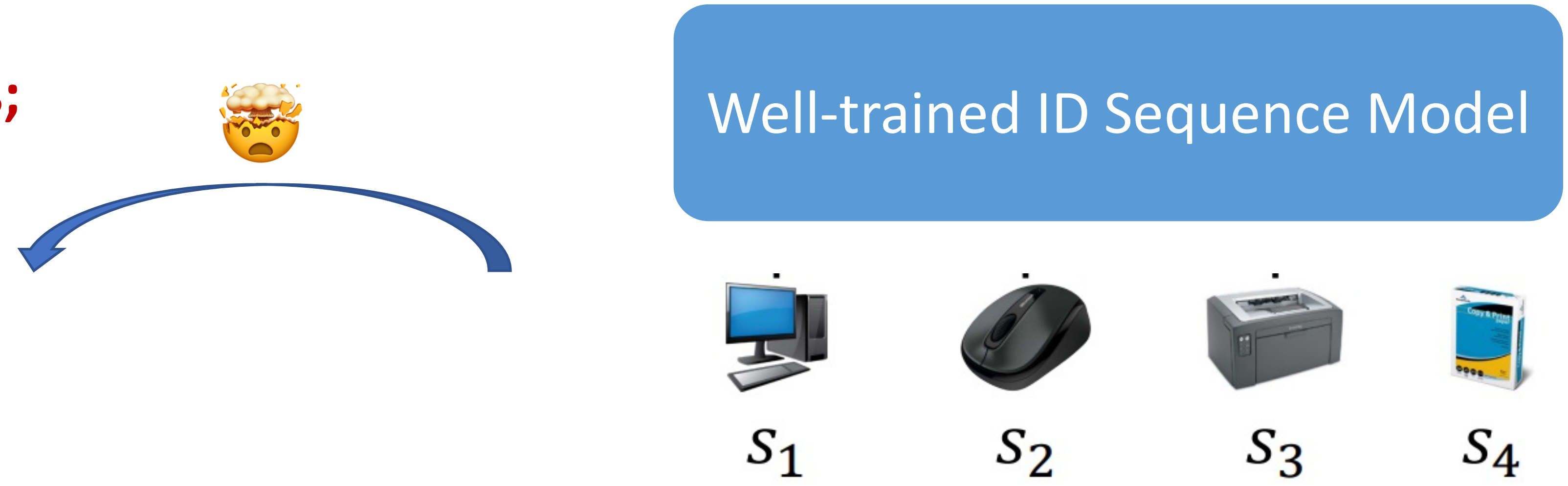


Kang and McAuley. Self-Attentive Sequential Recommendation. ICDM 2018.

Motivation

- Sequential Recommendation
  - Existing methods: explicit item ID modeling;

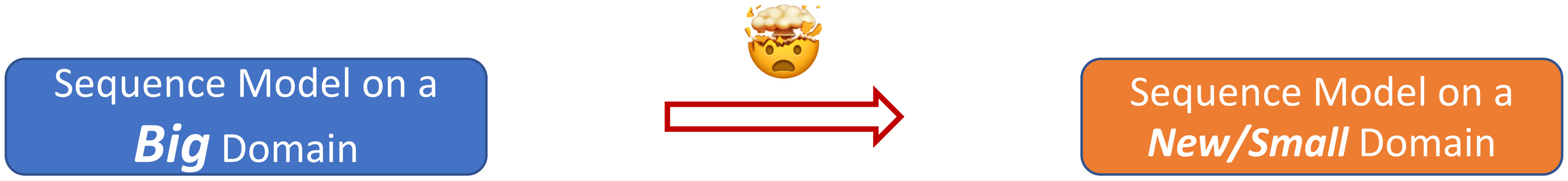
- Issues
  - **Cold-start items;**



New Item!

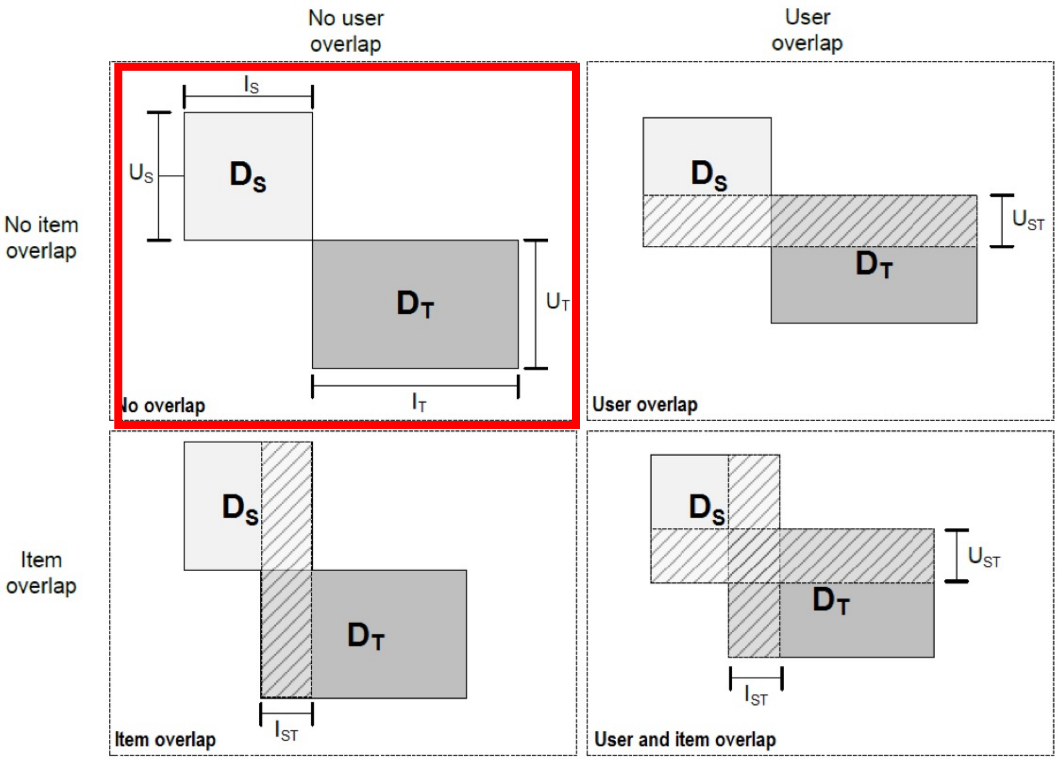
Motivation

- Sequential Recommendation
  - Existing methods: explicit item ID modeling;
- Issues
  - Cold-start items;
  - **New domains;**



Transfer Learning for RecSys

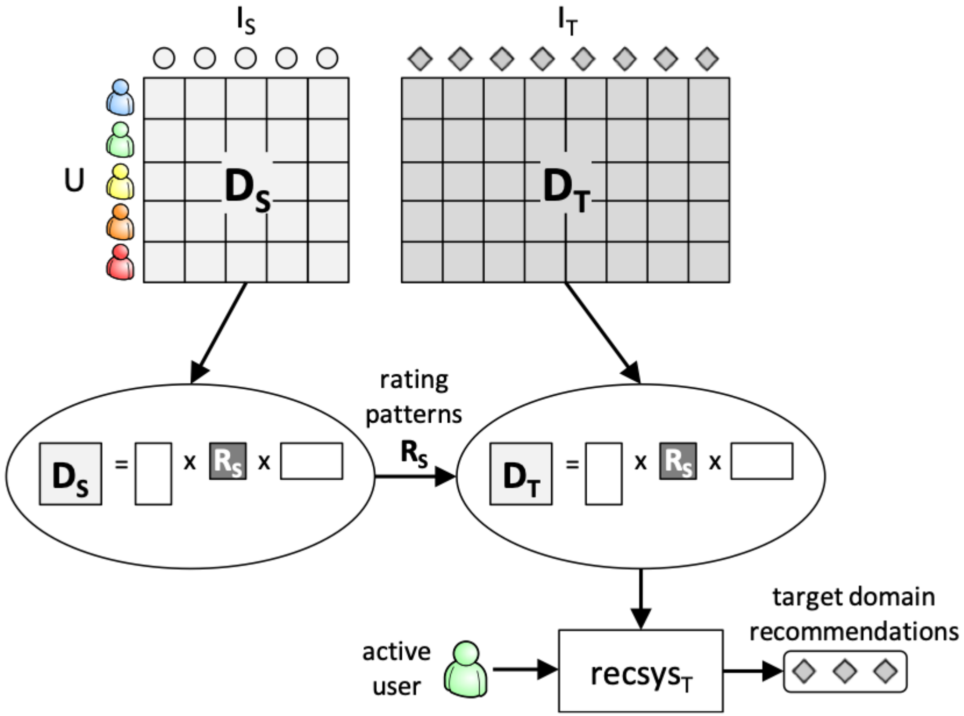
- It's hard to transfer recommendation models;
  - Need shared key anchors (e.g., users / items);



Cross-Domain Recommender Systems. **Recommender Systems Handbook 2015.**

Transfer Learning for RecSys

- It's hard to transfer recommendation models;
  - Need shared key anchors (e.g., rating patterns);



Can Movies and Books Collaborate? Cross-Domain Collaborative Filtering for Sparsity Reduction. **IJCAI 2009**

Motivation

- Foundation Models?

Language Models are Few-Shot Learners

We presented a 175 billion parameter language model which shows strong performance on many NLP tasks and benchmarks in the zero-shot, one-shot, and few-shot settings, in some cases nearly matching the performance of state-of-the-art fine-tuned systems, as well as generating high-quality samples and strong qualitative performance at tasks defined on-the-fly. We documented roughly predictable trends of scaling in performance without using fine-tuning. We also discussed the social impacts of this class of model. Despite many limitations and weaknesses, these results suggest that very large language models may be an important ingredient in the development of adaptable, general language systems.

Motivation

- Solution for small data: Transfer Learning
  - Foundation Models
    - Pre-training on large scale corpus;
    - Transfer to downstream tasks;
- Can we build foundation models for RecSys? 🤔

Motivation

- Solution for small data: Transfer Learning
  - Foundation Models
    - Pre-training on large scale corpus;
    - Transfer to downstream tasks;

- Can we build foundation models for RecSys? 🤔
  - ✅ **Same data format;**

s_1 s_2 s_3 s_4

Motivation

- Solution for small data: Transfer Learning
  - Foundation Models
    - Pre-training on large scale corpus;
    - Transfer to downstream tasks;

- Can we build foundation models for RecSys? 🤔
  - ✅ Same data format;
  - ✅ **Large corpus;** amazon Alibaba Group

Motivation

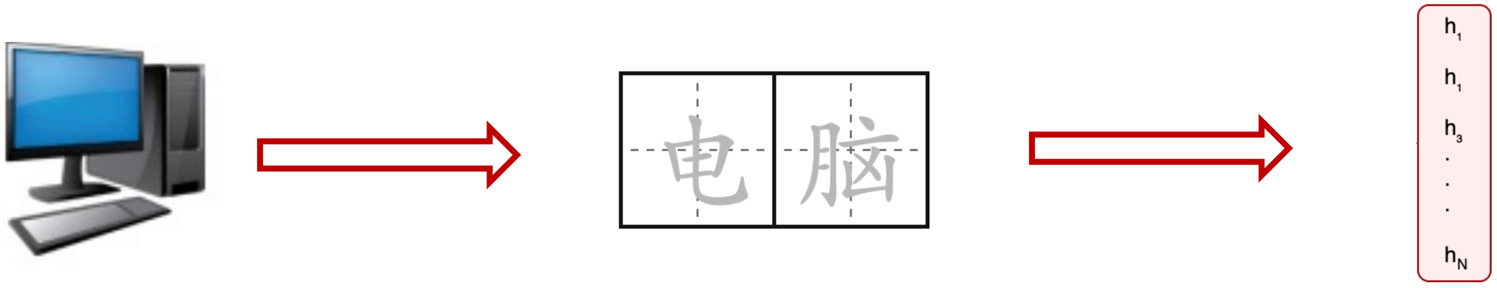
- Solution for small data: Transfer Learning
  - Foundation Models
    - Pre-training on large scale corpus;
    - Transfer to downstream tasks;

- Can we build foundation models for RecSys? 🤔
  - ✅ Same data format;
  - ✅ Large corpus;

  - ❌ **Different item ID dictionaries;**

Idea

- **Universal Sequence/Item Representations** for RecSys?



Item ID

Item Text

(natural language,

e.g., Title, Description, Brand, ……)

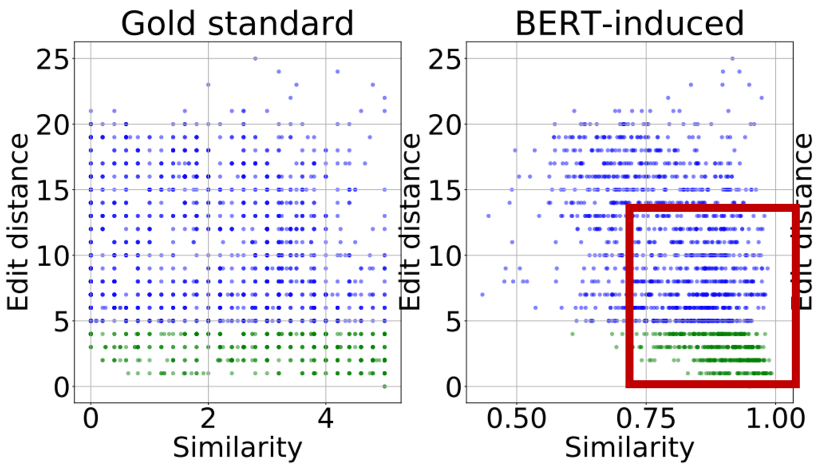
Item Emb

Sequence Emb

Idea

- Item text -> Transferable representations

- Challenges: PLM representations are not suitable for recommendation



Anisotropy[1]

* High similar scores between PLM reps.;

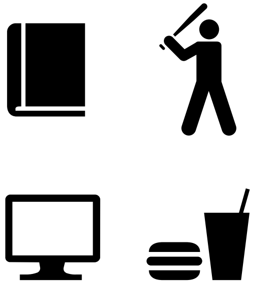
* Hard to distinguish;

1 On the Sentence Embeddings from Pre-trained Language Models. **EMNLP 2020.**

Idea

- Item text -> Transferable representations

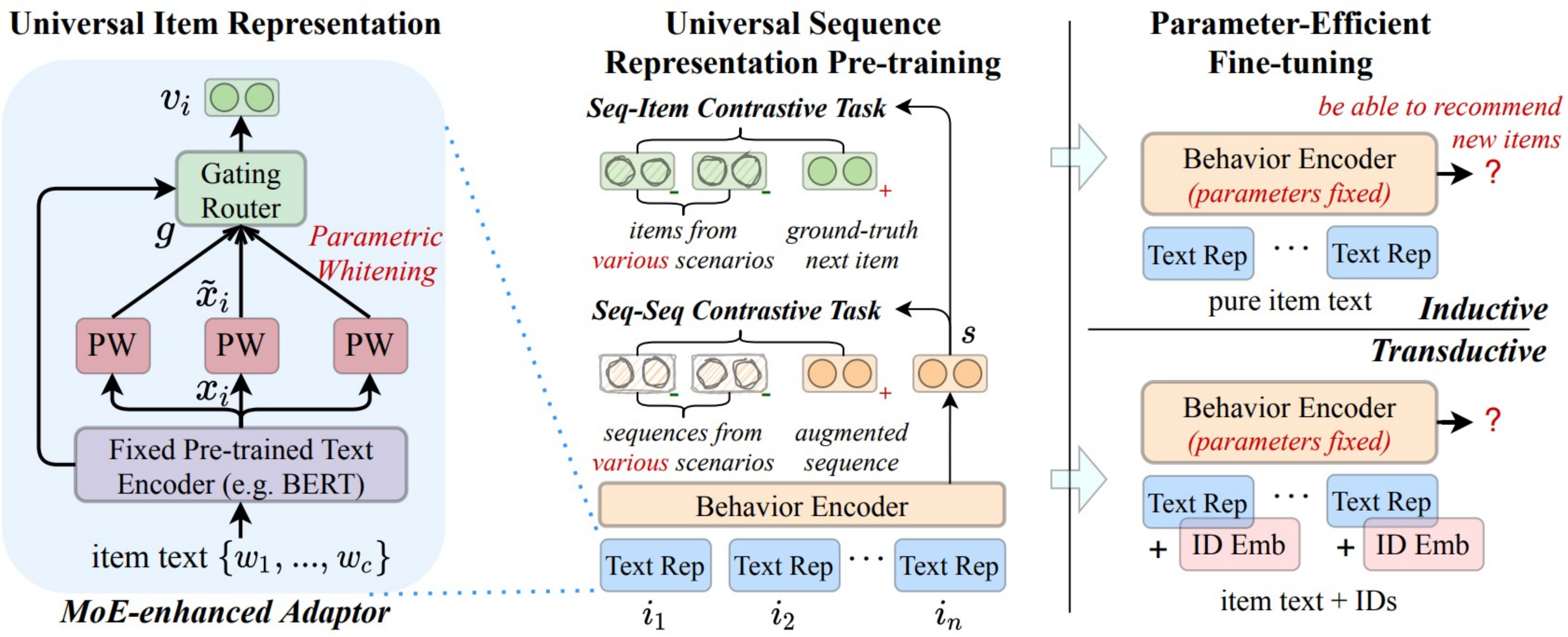
- Challenges: How to pre-train on multiple domains?



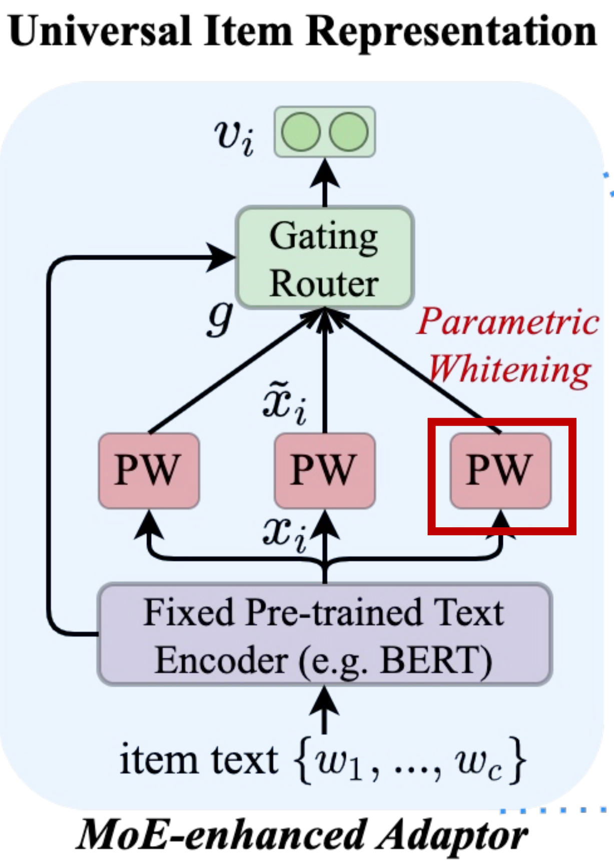
* Conflicts between domains while pre-training;

* How to adapt to new and unseen domains?

UniSRec, universal sequence representation learning approach



UniSRec

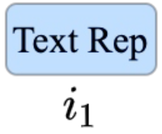


💡 Traditional whitening operations can not be adapted to new domains;

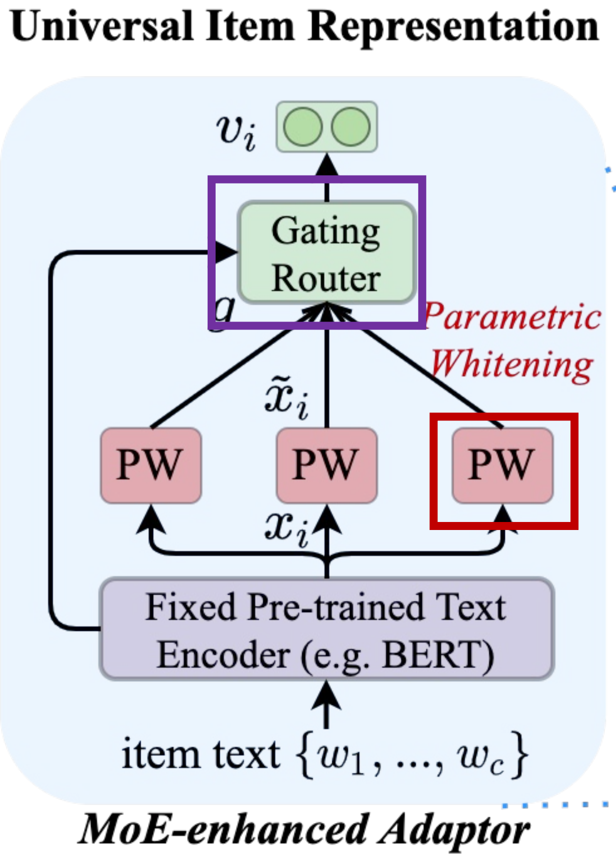
2. **Parametric Whitening** for semantic transformation & redusing anisotropy

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

1. PLM Encoding



UniSRec



(Mixture-of-Experts, MoE)

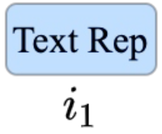
3. **MoE-enhanced Adaptor** for domain fusion & adaptation

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

2. **Parametric Whitening** for semantic transformation & redusing anisotropy

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

1. PLM Encoding

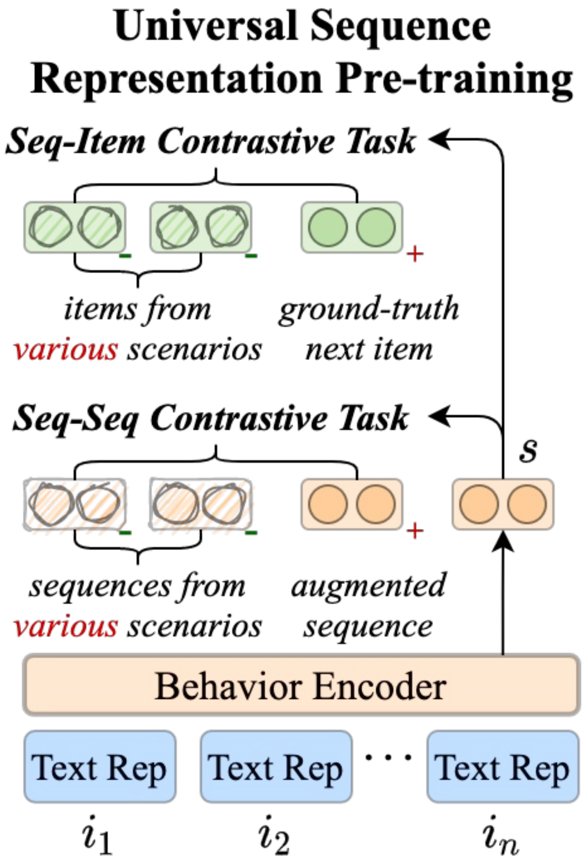


UniSRec

Pre-training on multi-domain sequences

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

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

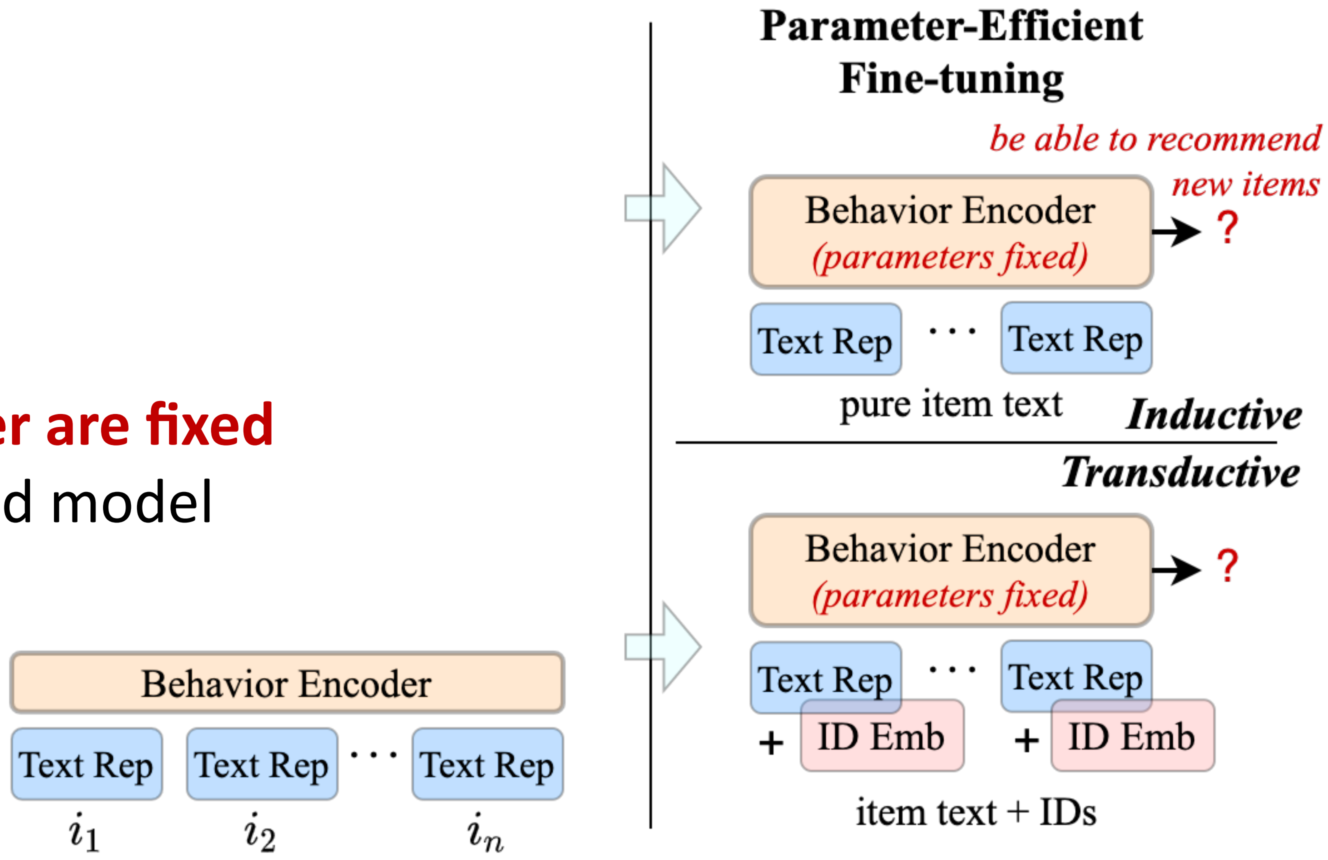


💡 **Negative samples from multiple domains** alleviate the seesaw phenomenon and capture their semantic correlation

Sequence model
e.g., Transformer encoder

UniSRec

Parameters in behavior encoder are fixed
while fine-tuning the pre-trained model



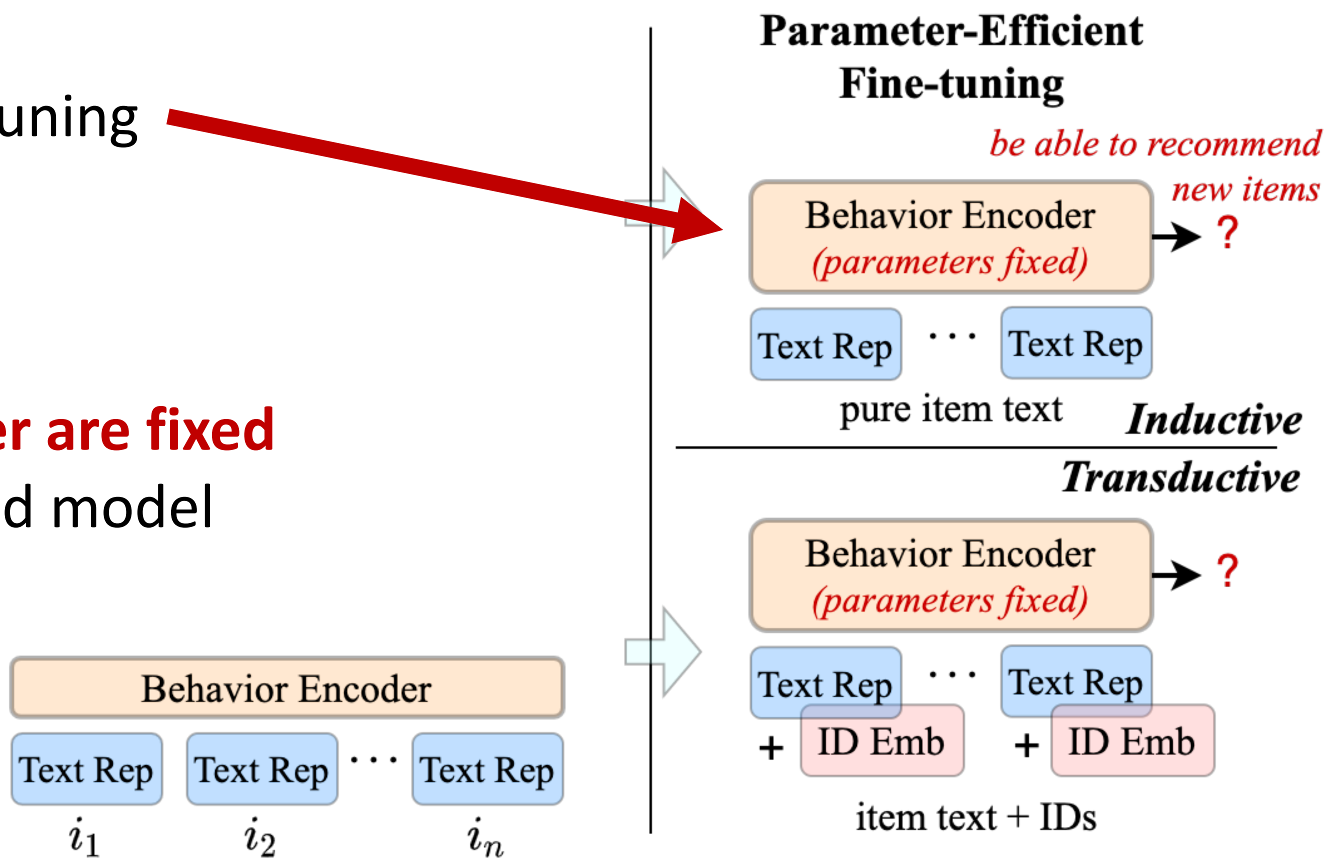
UniSRec

Inductive

Many new items, directly fine-tuning

Parameters in behavior encoder are fixed

while fine-tuning the pre-trained model

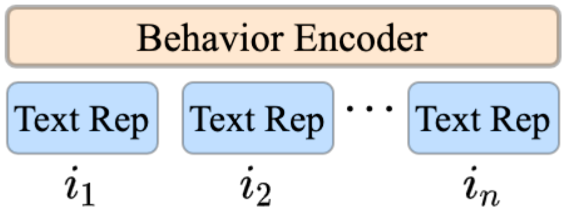


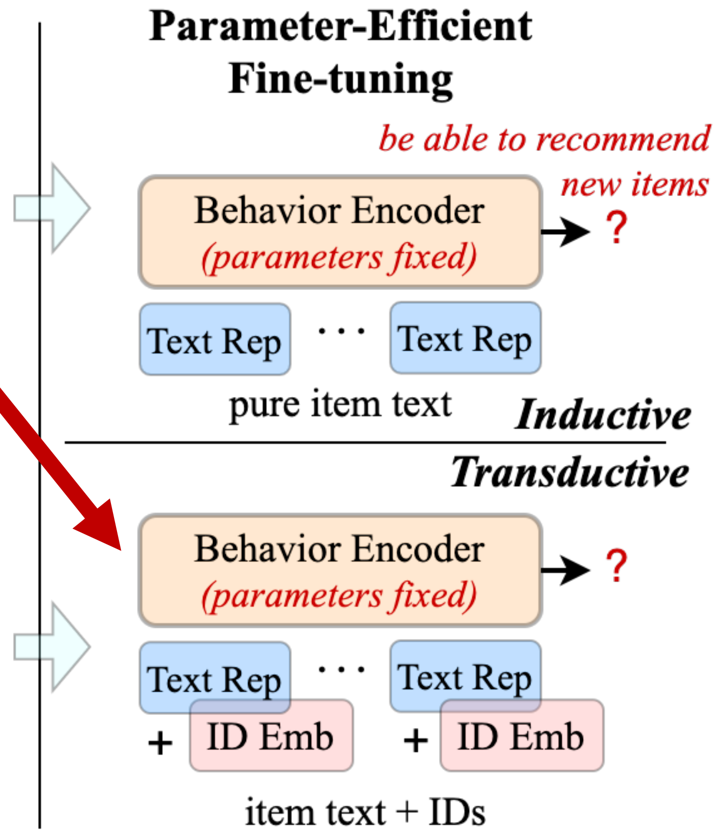
UniSRec

Transductive

Items are mostly seen while training,
ID embeddings are combined while fine-tuning

Parameters in behavior encoder are fixed
while fine-tuning the pre-trained model





Experiments – Setting

One UniSRec model

Pre-trained on

Amazon:
Food, CDs, Kindle, Movies, Home

Fine-tuned to

New domains
Scientific, Pantry, Instruments, Arts, Office

New platform
Online Retail in UK

No overlapping users or items b/w pre-training and downstream datasets

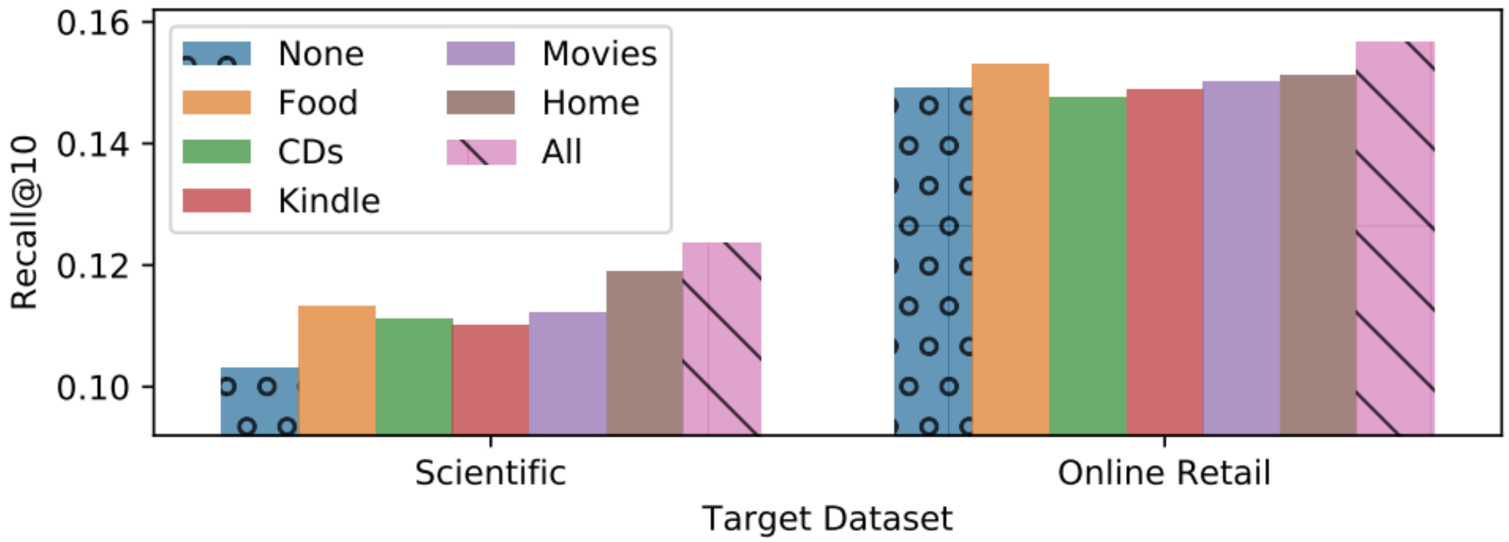
Benchmark for transferable rec. models

Experiments – Overall

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	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0586	0.0712	–
		Recall@50	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.2049*	**0.2347***	+15.62%
		NDCG@50	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0799	**0.0955***	+8.15%
	Office	Recall@10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.1013	**0.1280***	+11.79%
		NDCG@10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	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%

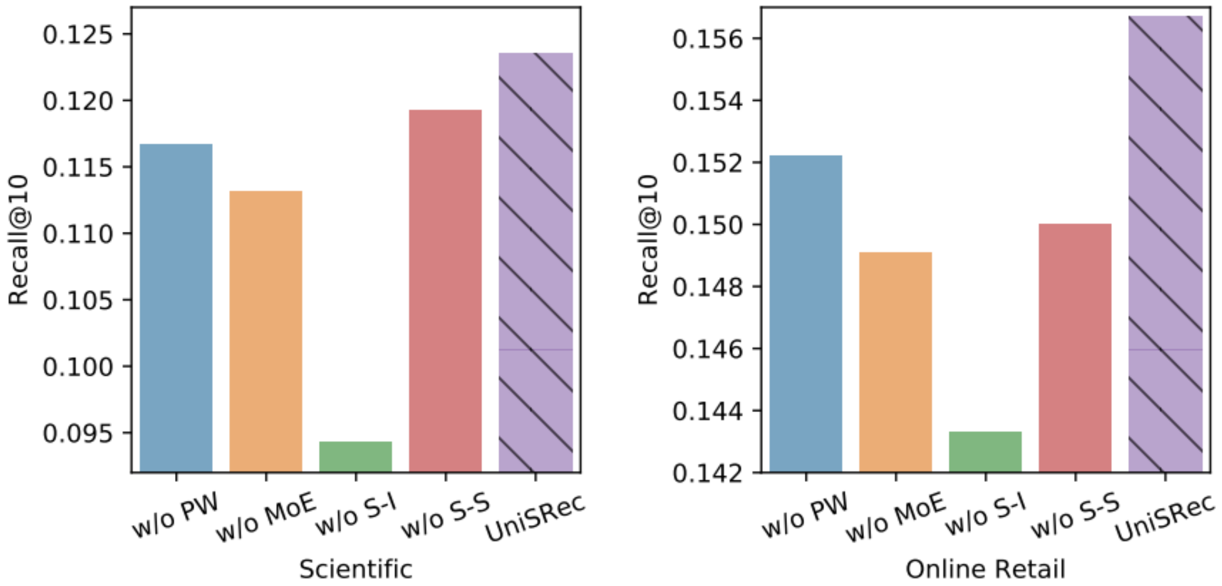
Improvements on x-domain & x-platform datasets

Experiments – Pre-trained datasets



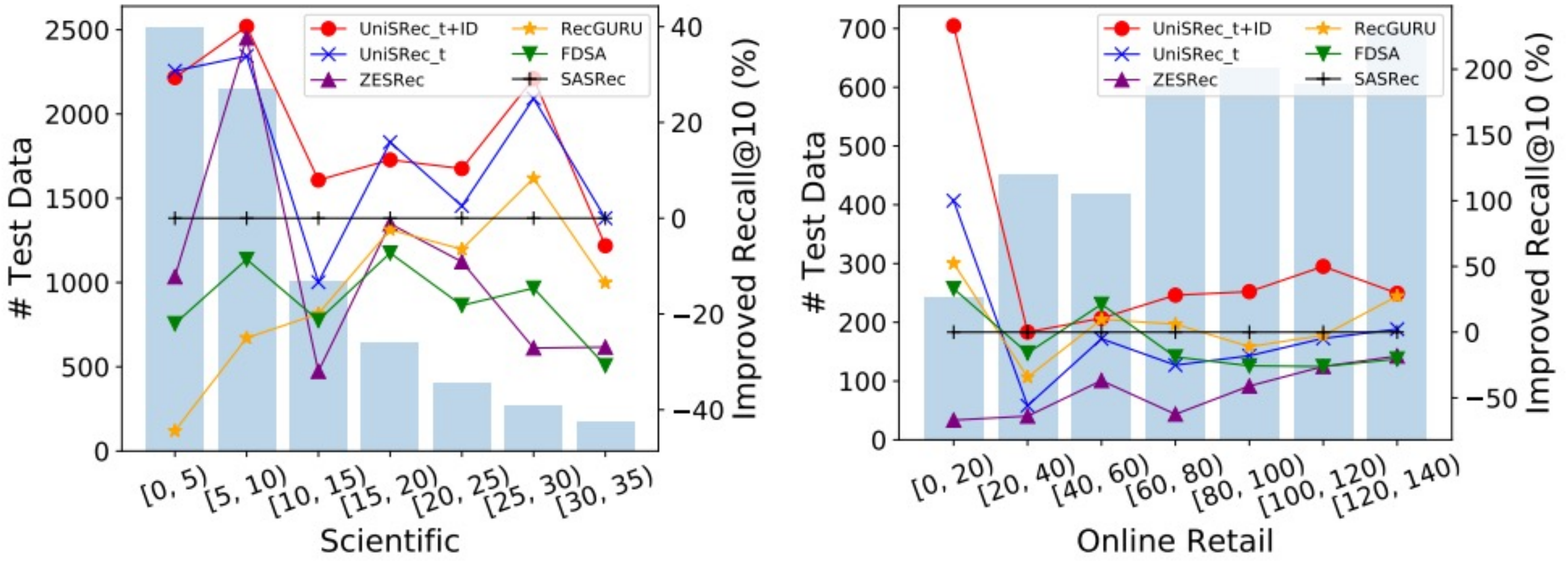
Pre-trained on multiple domains > pre-trained on single domain

Experiments – Ablation study



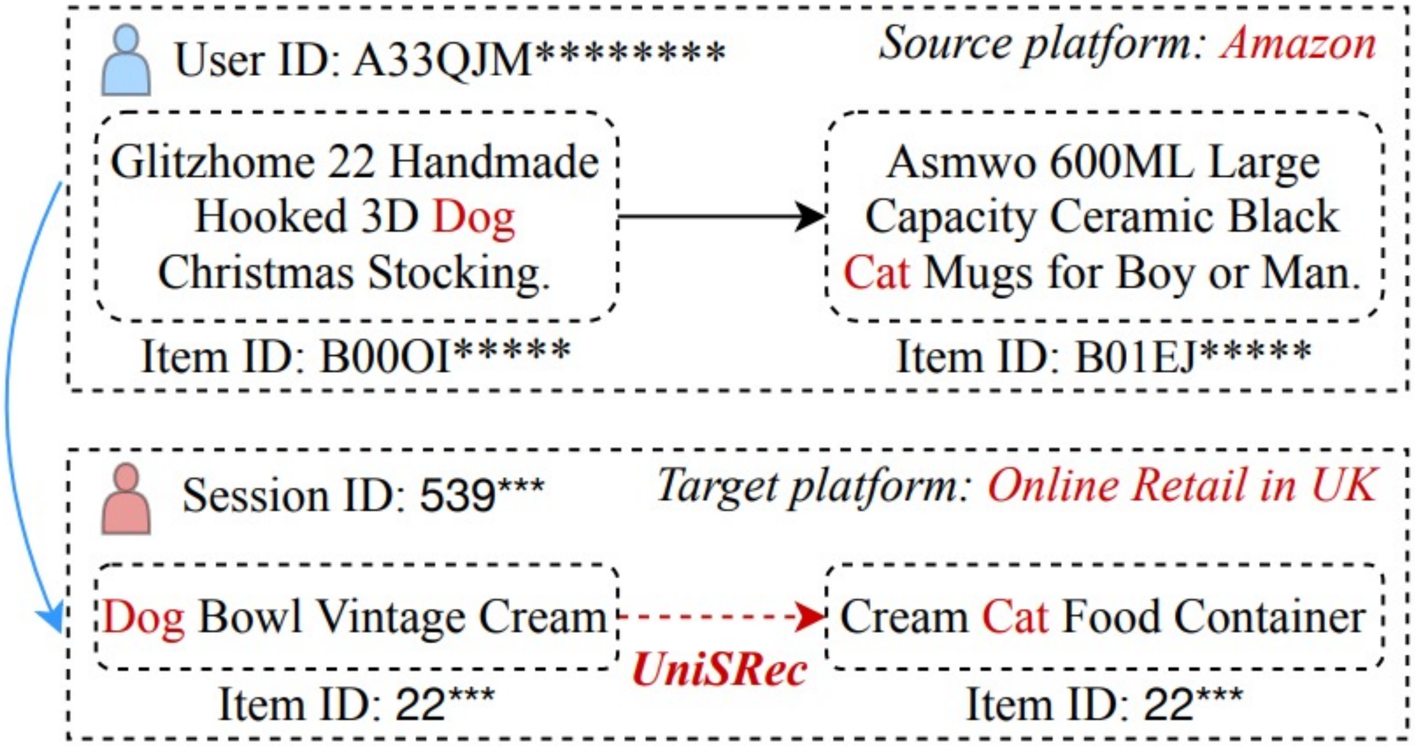
1) Parametric Whitening, 2) MoE-enhanced adaptor,
3) pre-training tasks all have beneficial effects.

Experiments – Long-tail items



Significant improvements over cold-start items

Experiments – Case study



Capture and transfer general semantic patterns from multiple source domains

AD for RecBole

Open-source one-station library for RecSys research

2k stars

38 watching

379 forks

Based on PyTorch

130+ models, 28 processed datasets

Extensions for GNN, Transformers, cross-domain, meta-learning, augmentation, debias, fairness, …

RecBole: Towards a unified, comprehensive and efficient framework for recommendation algorithms 81
WX Zhao, S Mu*, Y Hou*, Z Lin, Y Chen, X Pan, K Li, Y Lu, H Wang, ...
CIKM 2021, Resource Track

RecBole 2.0: Towards a More Up-to-Date Recommendation Library
WX Zhao, Y Hou*, X Pan*, C Yang, Z Zhang, Z Lin, J Zhang, S Bian, ...
CIKM 2022, Resource Track

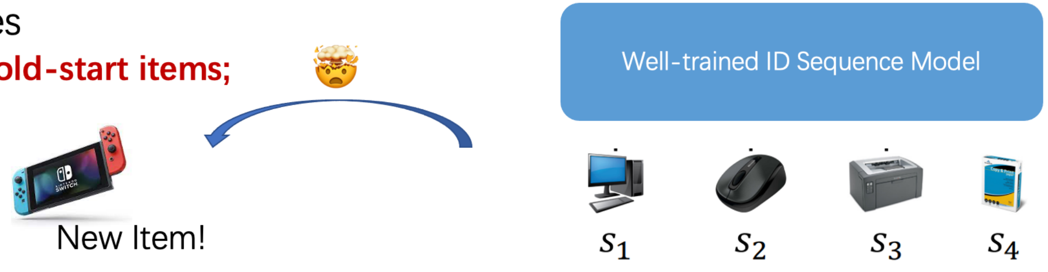
Conclusion & QA

presented by **Yupeng Hou**, **houyupeng@ruc.edu.cn**

https://github.com/RUCAIBox/UniSRec

All implemented by RecBole

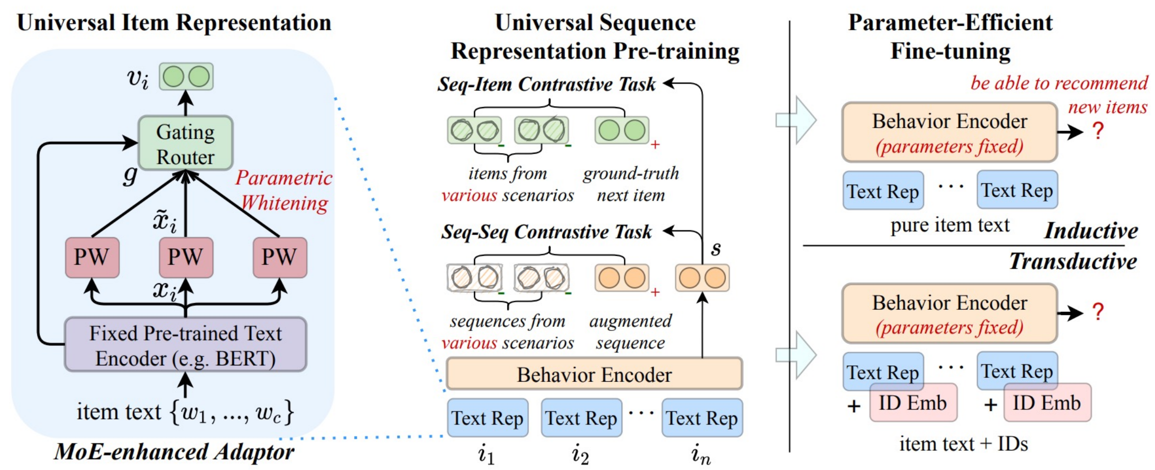
- Issues
  - **Cold-start items;**



  - **New domains;**

Sequence Model on a *Big* Domain → Sequence Model on a *New/Small* Domain

- **Universal Sequence/Item Representations** for RecSys?



One UniSRec model ← *Pre-trained on* **Amazon:** Food, CDs, Kindle, Movies, Home

Fine-tuned to

New domains
Scientific, Pantry, Instruments, Arts, Office

New platform
Online Retail in UK