Speaker: Fajie Yuan (PI of REPL Lab)

Time: 2023/09/23

Paradigm for Pre-training and Transfer Learning in Recommender Systems

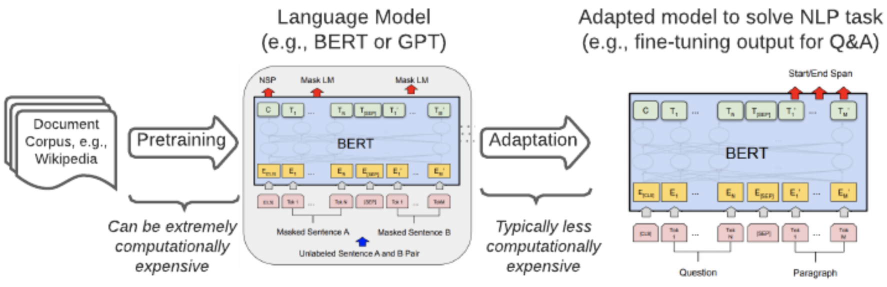
CONTENTS / 目录

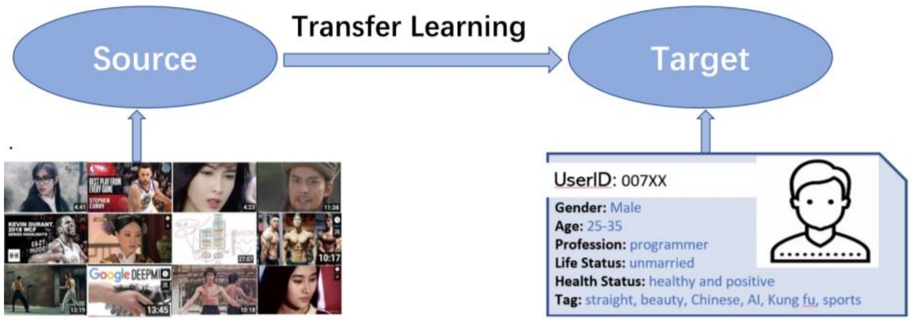
01

ID Overlapping-based Transfer

ID Overlapping-based Transfer

PeterRec (SIGIR2020)





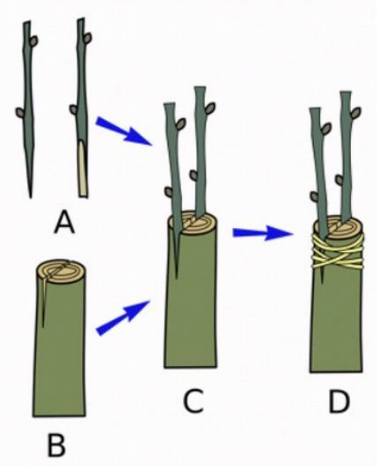
Source : (u, x^u), where $x^u = \{x_1^u, x_2^u, \dots x_n^u\}$,
where x_t^u denotes the t−th interacted item of user u

Target : (u, y) where y is the supervise label in the target dataset

Parameter-efficient transfer from sequential behaviors for user modeling and recommendation. SIGIR2020

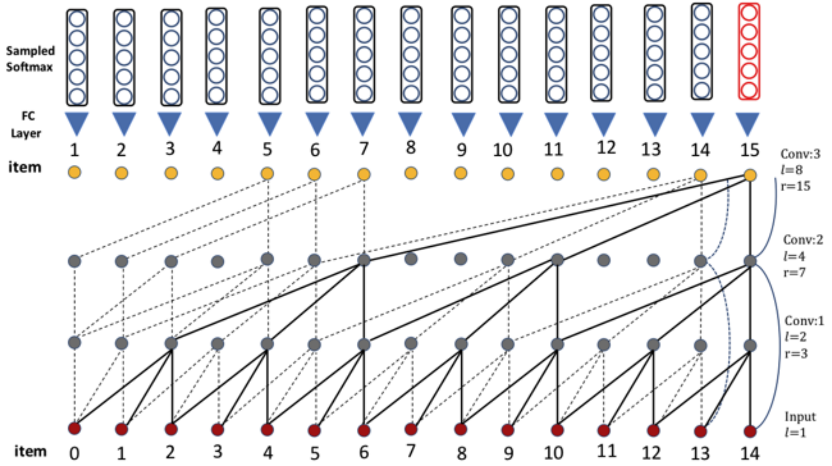
ID Overlapping-based Transfer

PeterRec (SIGIR2020)



Tree

A: branch of plum
B: Tree of peach
C: insertion
D: grow together



Pre-trained model

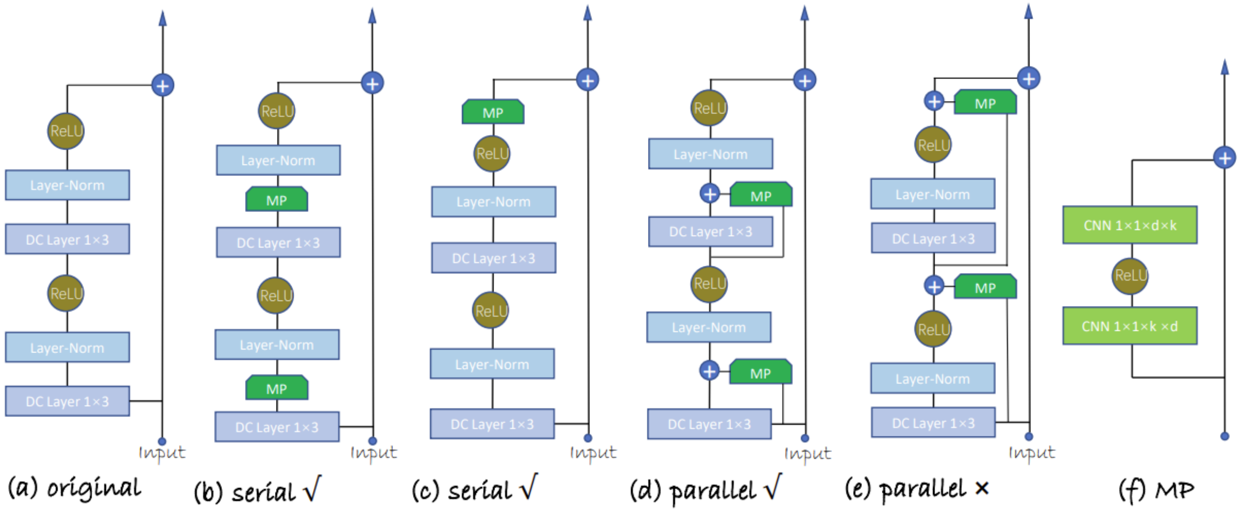
Pretrained model is treated as the peach Tree.

Parameter-efficient transfer from sequential behaviors for user modeling and recommendation. SIGIR2020

ID Overlapping-based Transfer

PeterRec (SIGIR2020)

How we do these insertion?



Parameter-efficient transfer from sequential behaviors for user modeling and recommendation. SIGIR2020

ID Overlapping-based Transfer

PeterRec (SIGIR2020)

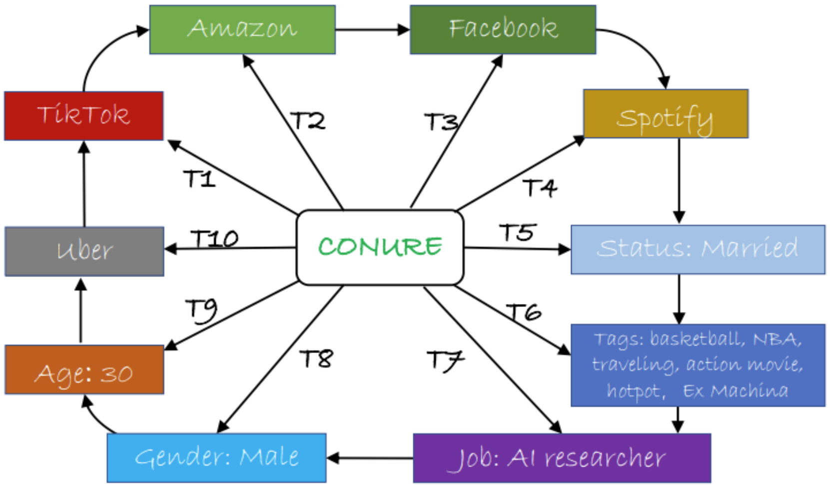
- The first work achieving transferable recommendation between domains
- Shared information is need for ID mapping between domains

ID Overlapping-based Transfer

Conure (SIGIR2021)

'一人一世界'

A person has different roles to play in different life scenes！But all these roles may have some commonalities, such as personalization, habits, preference.



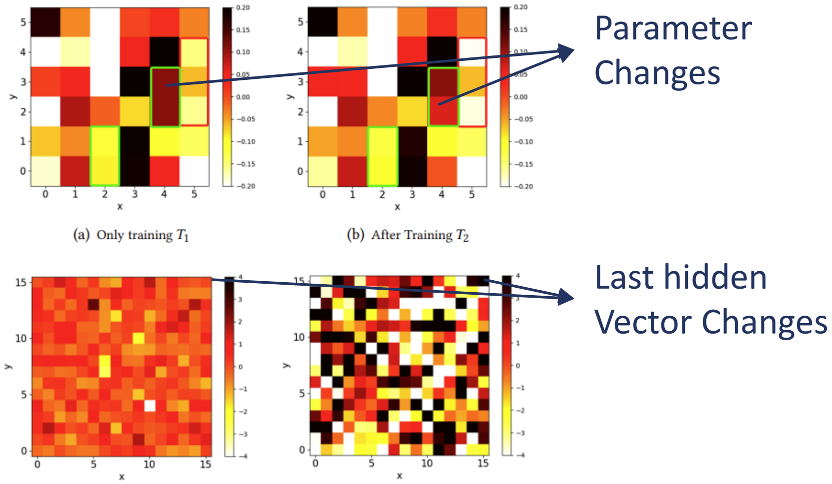
One Person, One Model, One World

One Person, One Model, One World: Learning Continual User Representation without Forgetting. SIGIR2021

ID Overlapping-based Transfer

Conure (SIGIR2021)

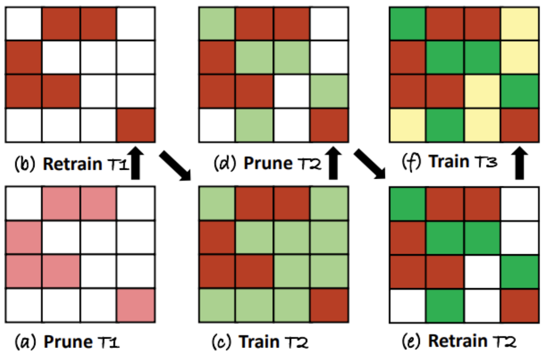
Catastrophic Forgetting !



(a) Only training T_1 (b) After Training T_2

(c) Only training T_1 (d) After training T_2

How Conure does:

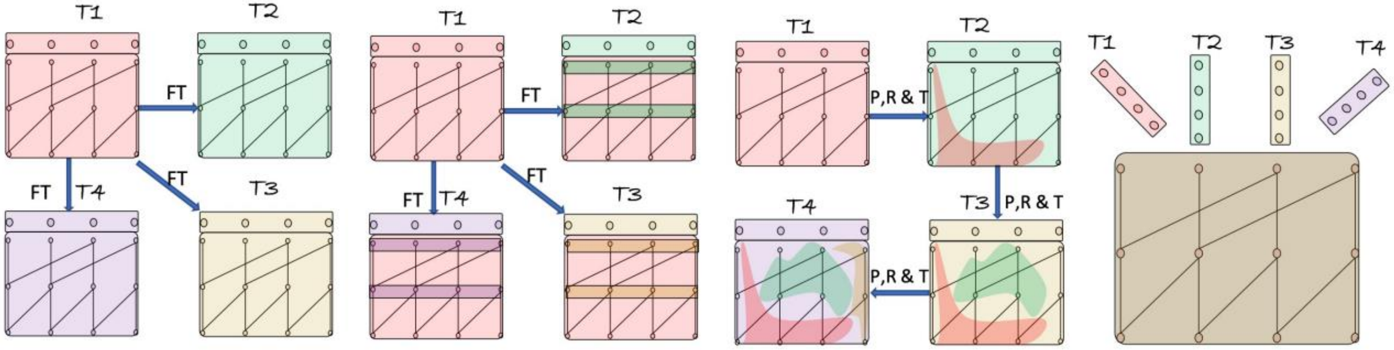


One Person, One Model, One World: Learning Continual User Representation without Forgetting. SIGIR2021

ID Overlapping-based Transfer

Conure (SIGIR2021)

Model Comparison:



(a)
Standard
Transfer Learning

(b)
PeterRec

(c)
Conure

(d)
multi-task learning
(MTL)

One Person, One Model, One World: Learning Continual User Representation without Forgetting. SIGIR2021

ID Overlapping-based Transfer

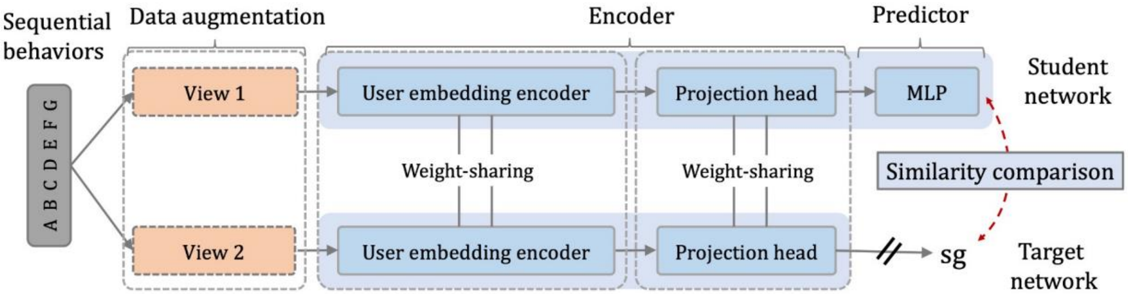
Conure (SIGIR2021)

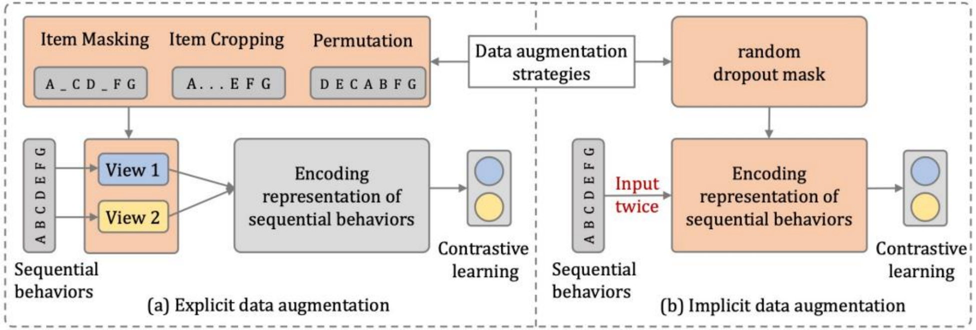
- The first work proposing lifelong learning in recommendation
- Shared information is still need

One Person, One Model, One World: Learning Continual User Representation without Forgetting

ID Overlapping-based Transfer

CLUE (ICDM2021)





Learning transferable user representations with sequential behaviors via contrastive pre-training. ICDM2021

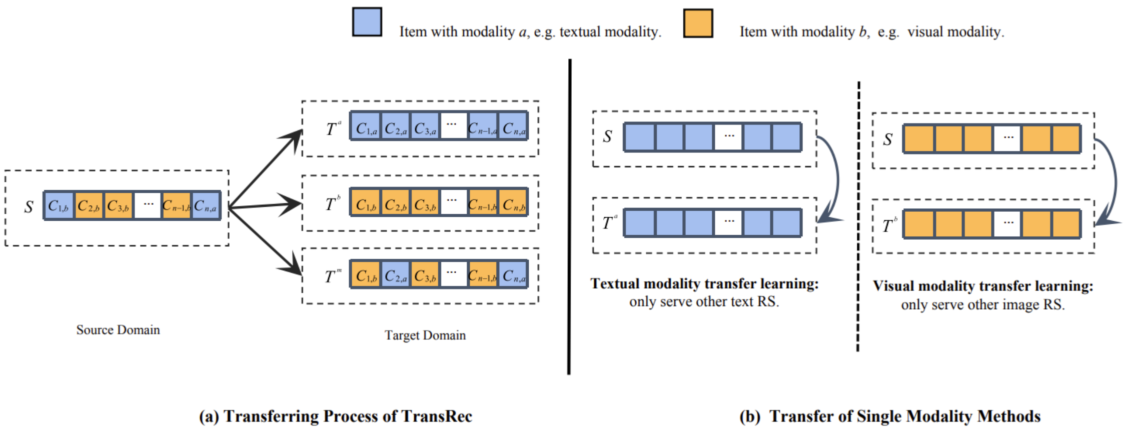
02

Modality-based Transfer

Modality-based Transfer

TransRec

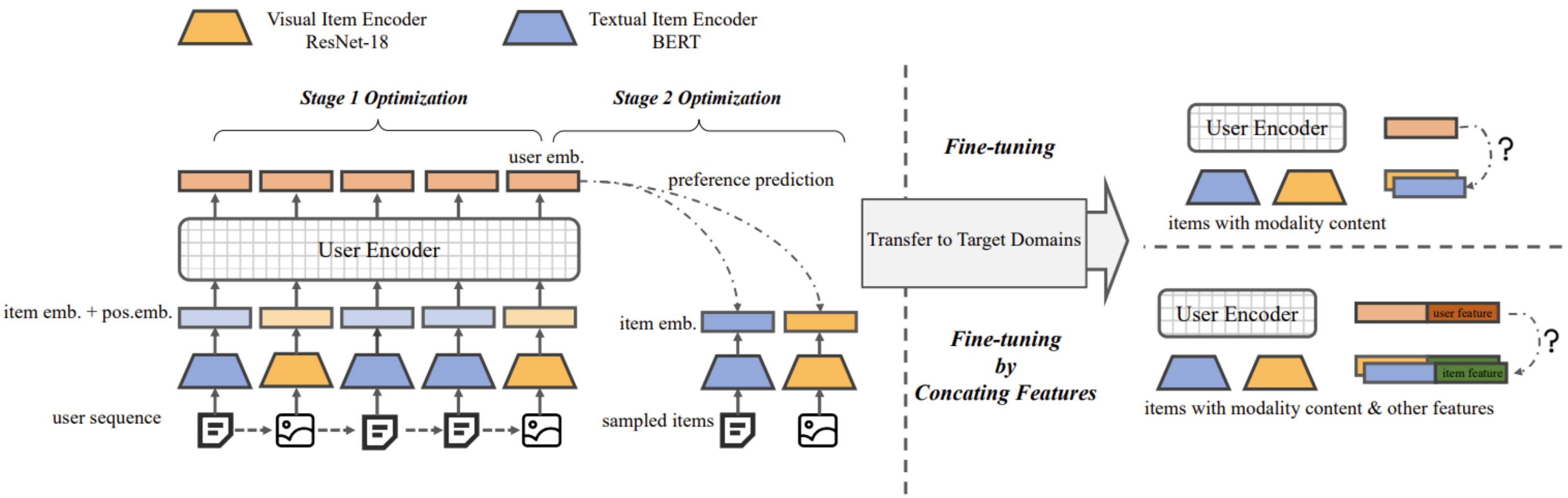
The first Recommender System regime enabling effective transfer across modalities & domains!



TransRec: Learning Transferable Recommendation from Mixture-of-Modality Feedback. 2022/06

Modality-based Transfer

TransRec



(a) Pre-training on Source Domain (b) Fine-tuning on Target Domains

TransRec: Learning Transferable Recommendation from Mixture-of-Modality Feedback. 2022/06

Modality-based Transfer

TransRec

Result:

Domain	Modality	Metric	IDRNN	IDCNN	IDRec	TFS	TransRec	Improv.
TN-mixed	Mixed	HR@5	0.0109	0.0112	0.0117	0.0249	**0.0285**	14.46%
		NDCG@5	0.0063	0.0067	0.0068	0.0160	**0.0177**	10.63%
		HR@10	0.0129	0.0195	0.0210	0.0428	**0.0478**	11.68%
		NDCG@10	0.0062	0.0094	0.0100	0.0213	**0.0239**	12.21%
TN-video	Image	HR@5	0.0134	0.0159	0.0153	0.0208	**0.0271**	30.29%
		NDCG@5	0.0093	0.0098	0.0092	0.0131	**0.0173**	30.06%
		HR@10	0.0201	0.0265	0.0267	0.0336	**0.0424**	26.19%
		NDCG@10	0.0114	0.0133	0.0125	0.0173	**0.0221**	27.75%
TN-text	Text	HR@5	0.0105	0.0123	0.0105	0.0303	**0.0358**	18.15%
		NDCG@5	0.0063	0.0078	0.0062	0.0192	**0.0227**	18.23%
		HR@10	0.0189	0.0220	0.0192	0.0500	**0.0597**	19.40%
		NDCG@10	0.0089	0.0109	0.0090	0.0255	**0.0303**	18.82%
DouYin	Image	HR@5	0.0059	0.0057	0.0023	0.0115	**0.0146**	26.96%
		NDCG@5	0.0037	0.0035	0.0014	0.0073	**0.0090**	23.29%
		HR@10	0.0096	0.0100	0.0035	0.0205	**0.0259**	26.34%
		NDCG@10	0.0049	0.0049	0.0018	0.0101	**0.0126**	24.75%

a. TransRec performs consistently better than its training-from-scratch version, i.e., TFS.
b. TransRec performs better than ID-based methods as well.

TransRec: Learning Transferable Recommendation from Mixture-of-Modality Feedback

Modality-based Transfer

MoRec (SIGIR2023)

MoRec vs. IDRec

IDRec's weaknesses :

a. Fails when users and items have few interactions, e.g., the cold-start setting.
b. Pre-trained IDRec is not transferable across platforms given that userIDs and itemIDs are in general not shareable in practice.
c. Pure IDRec cannot benefit from the technical advances in other communities NLP and CV.

MoRec's potential:

a. MoRec is comparable to or even surpass IDRec in no-cold setting.
b. MoRec build connections for RS and other communities and inherit their latest advances.
c. Pre-trained MoRec can transfer across domains even without shared information.

Modality-based Transfer

MoRec (SIGIR2023)

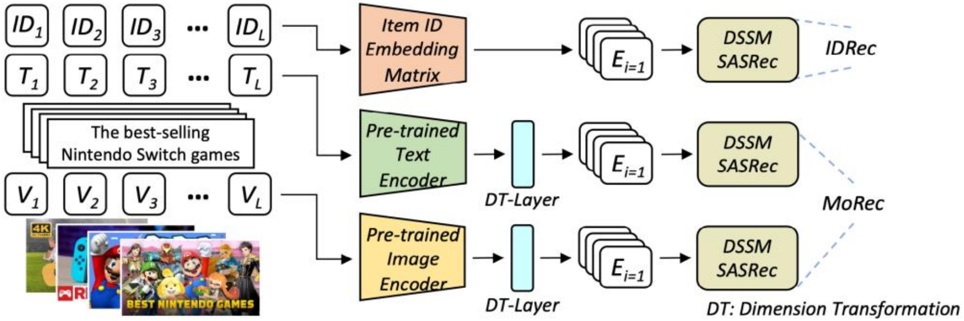


Illustration of IDRec vs MoRec. The only difference is the item encoder.

- IDRec uses an item ID embedding matrix.
- MoRec uses the pre-trained modality encoder.

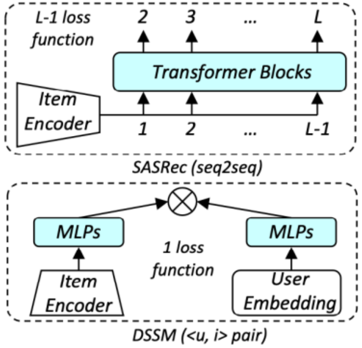


Illustration of DSSM and SASRec.

Modality-based Transfer

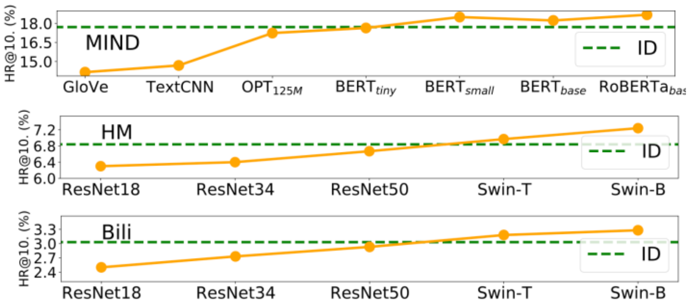
MoRec (SIGIR2023)

Diverse pure modal-based dataset

(a) Item cases on ImageNet1K. (b) Item cases on HM. (c) Item cases on Bili.

Dataset	Metrics	DSSM			SASRec				Improv.
		IDRec	BERT$_{base}$	RoBERTa$_{base}$	IDRec	BERT$_{small}$	BERT$_{base}$	RoBERTa$_{base}$	
MIND	HR@10	**3.58**	2.68	3.07	17.71	18.50	18.23	**18.68**	+5.48%
	NDCG@10	**1.69**	1.21	1.35	9.52	9.94	9.73	**10.02**	+5.25%
		IDRec	ResNet50	Swin-T	IDRec	ResNet50	Swin-T	Swin-B	
HM	HR@10	**4.93**	1.49	1.87	6.84	6.67	6.97	**7.24**	+5.85%
	NDCG@10	**2.93**	0.75	0.94	**4.01**	3.56	3.80	3.98	-0.75%
Bili	HR@10	**1.14**	0.38	0.57	3.03	2.93	3.18	**3.28**	+8.25%
	NDCG@10	**0.56**	0.18	0.27	1.63	1.45	1.59	**1.66**	+1.84%

MoRec vs IDRec (Regular Setting)



Accuracy with different pre-trained ME in MoRec.

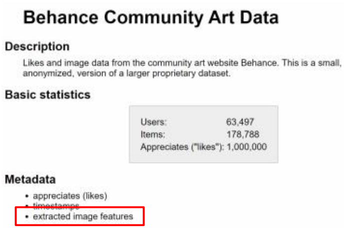
Modality-based Transfer

NineRec

Existing datasets pose a major obstacle for Transferable Recommender Systems!

Amazon

GEST



Behance

Title:
Mike Tomlin: Steelers 'accept responsibility' for role in brawl with Browns

Category:
Sport

...

MIND

E-commerce

Modality-only TransRec is hard to learn on E-commerce dataset where price matters a lot

Single Scenario

Interaction from a single scenario suffer from semantic insufficiency for Transfer learning

Pre-extracted Modality

Large representation gap between RS and CV&NLP is difficult to remove

Single Modality

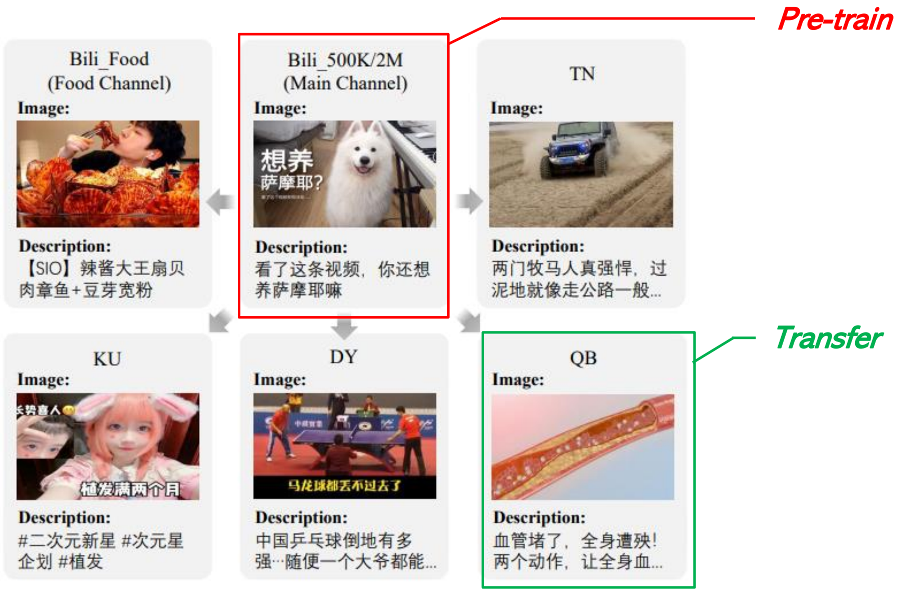
Single textual or visual information struggles to reflect the user preference

Modality-based Transfer

NineRec

A large-scale benchmark dataset for exploring MoRec' transferability between non-overlapping domains

- ***Dataset scale:***
  - 1 source: #User: 2M, #Item: 140k
  - 9 targets: #User: 2k-20k, #Item: 1k-8k
- ***No user overlap across targets***
- ***Raw text and image***
- ***Item from video platform:***
  - Interact mainly depend on content itself
  - Multiple targets across domain & platform
  - Diverse & semantically rich topics



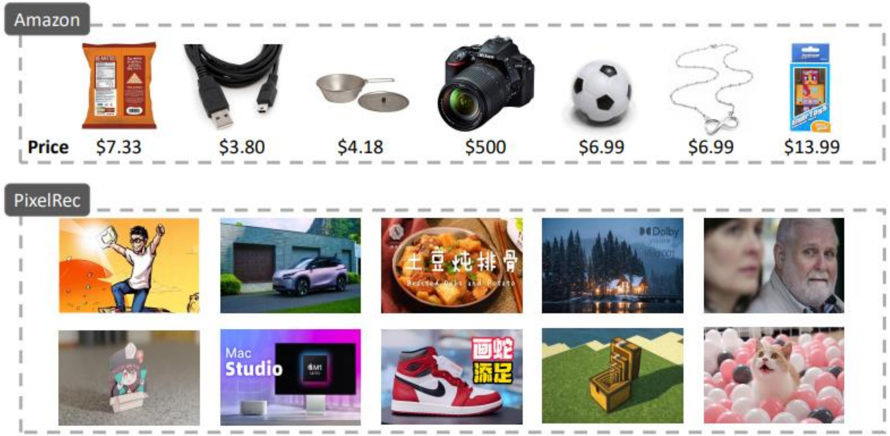
NineRec: A Benchmark Dataset Suite for Evaluating Transferable Recommendation. 2023/09

Modality-based Transfer

PixelRec

A large-scale benchmark dataset for pure image-centric MoRec

- ***Dataset scale:***
    - 200 million user-image interactions
    - 30 million users
    - 400,000 high-quality cover images
- ***High-resolution raw image***
- ***Rich features***
- ***Diverse content topics***

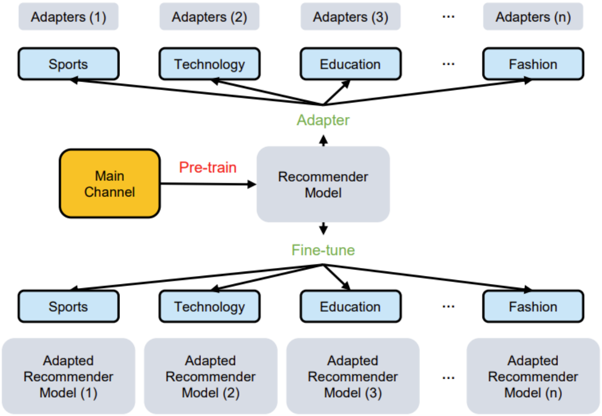


PixelRec: An Image Dataset for Studying Recommendations from Raw Pixels. 2023/09

Modality-based Transfer

Adapter-based TransRec

How to transfer in an efficient manner?



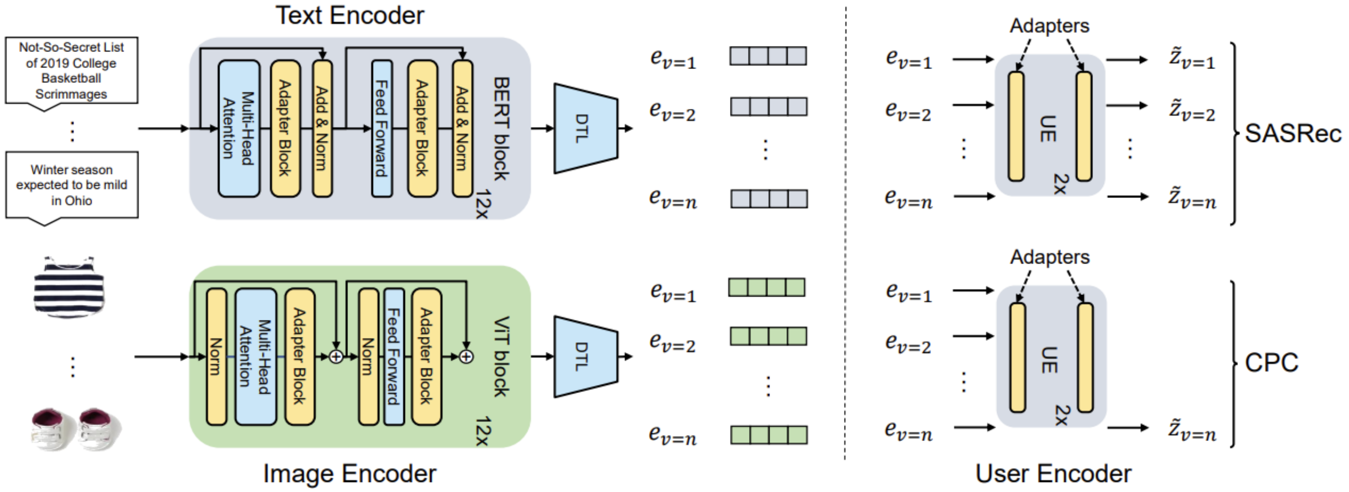
Modal-based transfer for downstream domains may heavy cost!

Exploring Adapter-based Transfer Learning for Recommender Systems: Empirical Studies and Practical Insights. 2023/05

Modality-based Transfer

Adapter-based TransRec

Only fine-tune Adapter networks when do transfer



Modality-based Transfer

Adapter-based TransRec

Text Scenario:

Image Scenario:

Datasets	Architecture	Metrics	FTA	AdaT	Difference
MIND->Adressa	SASRec+BERT (**Text**)	HR@10	**32.83**	32.52	-0.94%
		NDCG@10	17.33	**17.44**	+0.63%
	CPC+BERT (**Text**)	HR@10	29.56	**30.07**	+1.69%
		NDCG@10	15.81	**16.12**	+1.92%
		Trainable Parameters	100%	2.23%	-97.77%
	SASRec+RoBERTa (**Text**)	HR@10	32.02	**33.14**	+3.38%
		NDCG@10	16.95	**17.54**	+3.36%
	CPC+RoBERTa (**Text**)	HR@10	29.90	**30.64**	+2.42%
		NDCG@10	15.86	**16.20**	+2.10%
		Trainable Parameters	100%	1.95%	-98.05%
H&M->Amazon	SASRec+ViT (**Image**)	HR@10	**29.00**	27.66	-4.59%
		NDCG@10	**25.61**	24.36	-4.88%
	CPC+ViT (**Image**)	HR@10	**26.56**	25.29	-4.78%
		NDCG@10	**22.09**	21.49	-2.72%
		Trainable Parameters	100%	2.82%	-97.18%
	SASRec+MAE (**Image**)	HR@10	**28.10**	25.67	-8.61%
		NDCG@10	**22.92**	21.99	-4.05%
	CPC+MAE (**Image**)	HR@10	**27.50**	25.18	-8.44%
		NDCG@10	**23.51**	21.83	-7.14%
		Trainable Parameters	100%	2.82%	-97.18%

Comparable results but only 3% parameters fine-tuned

Still worse than fine-tuning all parameters

Exploring Adapter-based Transfer Learning for Recommender Systems: Empirical Studies and Practical Insights. 2023/05

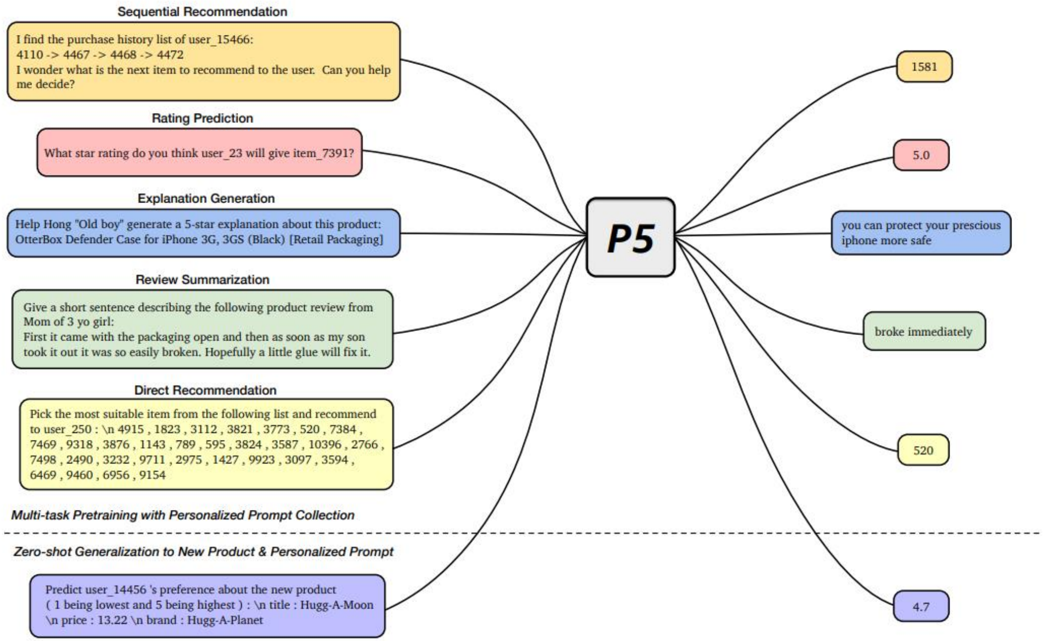
03

LLM-based Transfer

LLM-based Transfer

P5 (RecSys 2022)

The first task-agnostic pre-training framework in Recommender System

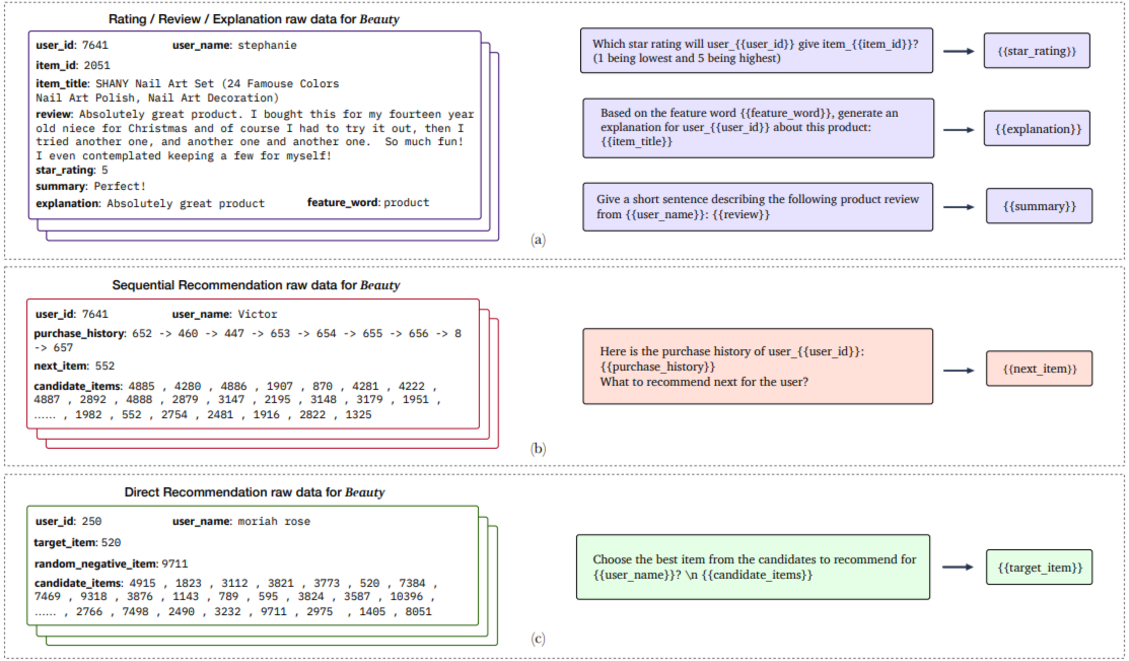


Recommendation as Language Processing (RLP): A Unified Pretrain, Personalized Prompt & Predict Paradigm (P5). Recsys2022

LLM-based Transfer

P5 (RecSys 2022)

ID-based
Prompt Engineering:

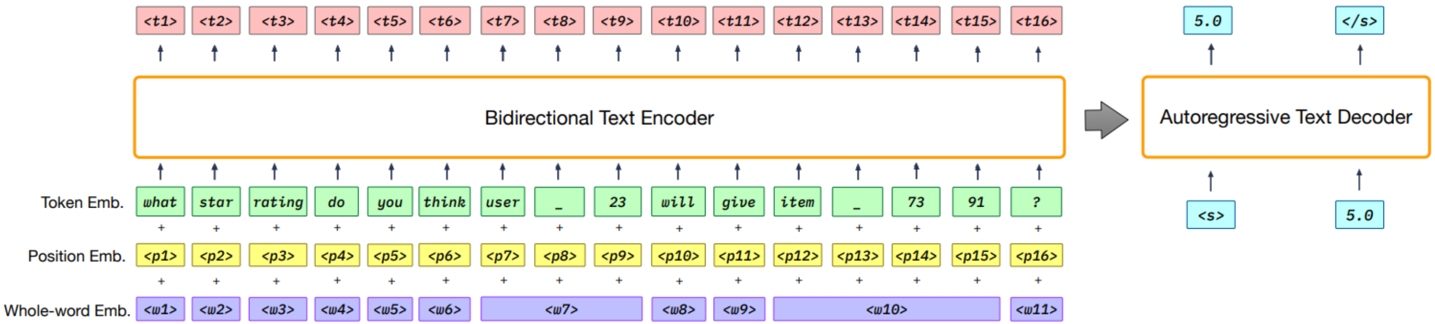


Recommendation as Language Processing (RLP): A Unified Pretrain, Personalized Prompt & Predict Paradigm (P5)

LLM-based Transfer

P5 (RecSys 2022)

P5 architecture:

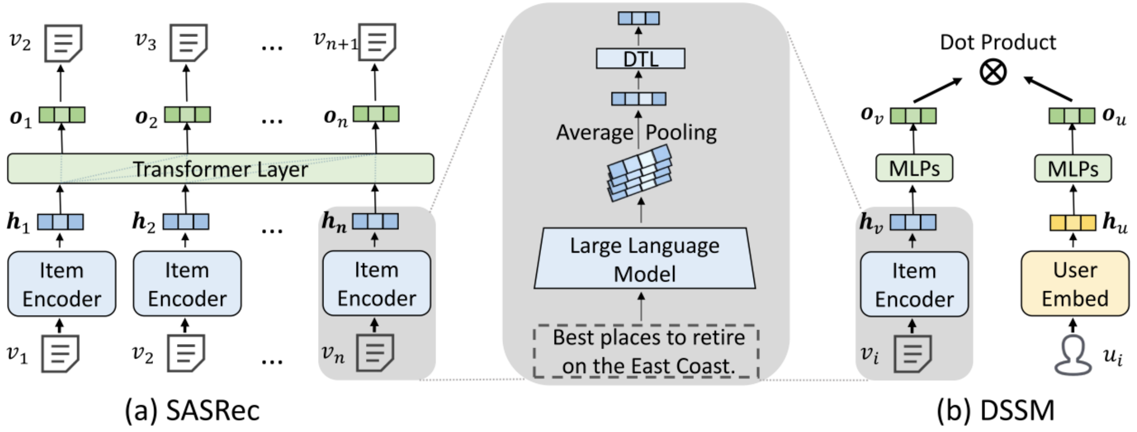


Recommendation as Language Processing (RLP): A Unified Pretrain, Personalized Prompt & Predict Paradigm (P5)

LLM-based Transfer

LLM4Rec

Does LLM enable emergent ability for Recommender System?

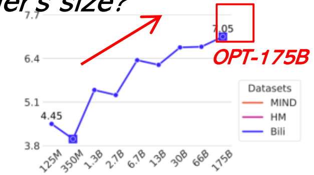


Exploring the Upper Limits of Text-Based Collaborative Filtering Using Large Language Models: Discoveries and Insights

LLM-based Transfer

LLM4Rec

Q(i): Does RS performance respond to the continuous increase in the item encoder's size?



Q(ii): Can the 175B parameter LLM achieve universal text representation?

Fine-tuned 125M > Frozen 175B

Q(iv): How close is the LLMs to a universal recommendation model?

Model	MIND	HM	QB
Random	0.02	0.01	0.18
$175B_{zero}$	0.13	0.39	4.30
$175B_{train}$	20.24	11.11	29.90

Q(iii): Can the 175B parameter LLM easily beat the simplest ID embedding based models (IDCF)?

Data	SASRec			DSSM		
	IDCF	$175B^{FR}$	$66B^{FT}$	ID	$175B^{FR}$	$66B^{FT}$
MIND	20.05	20.24	**21.07**	**3.99**	2.83	3.27
HM	12.02	11.24	**13.29**	**6.79**	2.09	2.35
Bili	7.01	6.88	**8.15**	**2.27**	2.00	2.01

Q(v): Will recent prompt engineering based RS utilizing ChatGPT challenge MoRec with LLMs?

Data	Task 1-HR@1				Task 2-HR@10			
	Random	ChatGPT	TCF_{175B}^{FR}	TCF_{66B}^{FT}	Random	ChatGPT	TCF_{175B}^{FR}	TCF_{66B}^{FT}
MIND	25.00	25.68	96.48	96.58	10.00	9.86	97.07	97.9
HM	25.00	29.59	88.18	90.63	10.00	12.21	83.79	90.33
Bili	25.00	24.51	77.64	81.05	10.00	8.50	70.80	73.34

better →

Exploring the Upper Limits of Text-Based Collaborative Filtering Using Large Language Models: Discoveries and Insights

LLM-based Transfer

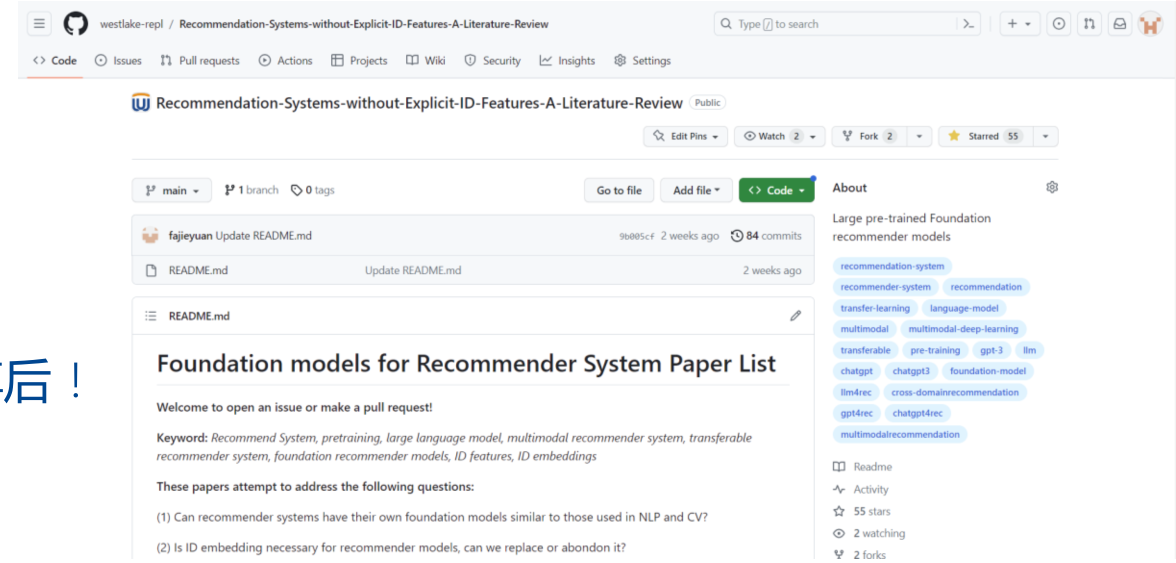
Find our GitHub:

westlake-repl

17 followers China westlake-repl@outlook.com

Updated collection about：

Pre-training and transfer learning in Recommender Systems



招实习！科研助理！博后！
支撑大模型训练！

THANKS

Fajie Yuan

2023/09/23