Where to Go Next for Recommender Systems? ID- vs. Modality-based Recommender Models Revisited

Zheng Yuan[1], Fajie Yuan[1], Yu Song[1], Youhua Li[1], Junchen Fu[1],
Fei Yang[2], Yunzhu Pan[1], Yongxin Ni[1]

[1]Westlake University; [2]Zhejiang Lab

西湖大學
WESTLAKE UNIVERSITY

Overview

Background and Motivation

Framework and Training Details

Experiments and Findings

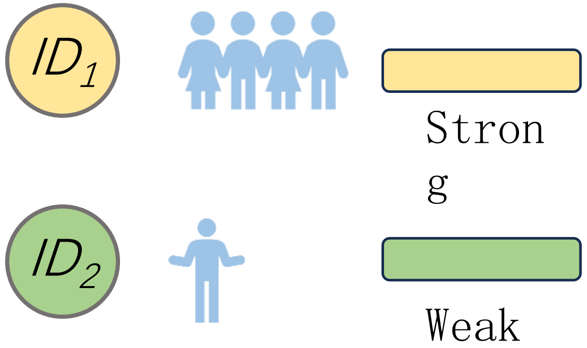
Future Works

Background and Motivation

Key weaknesses of ID-based Recsys (IDRec)

Background and Motivation

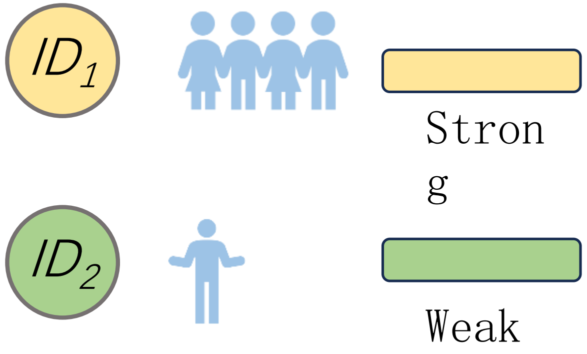
Key weaknesses of ID-based Recsys (IDRec)



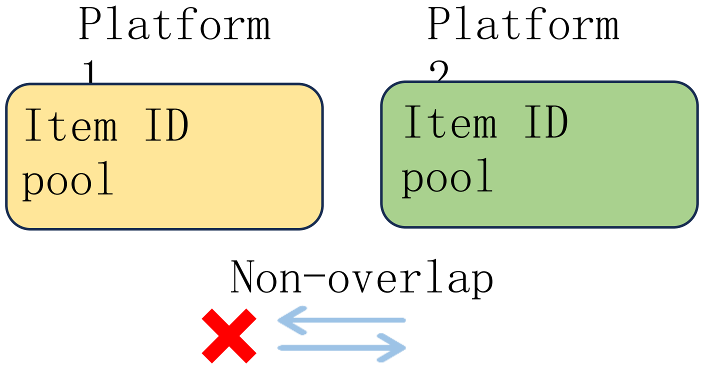
(1) Cold-start
setting

Background and Motivation

Key weaknesses of ID-based Recsys (IDRec)



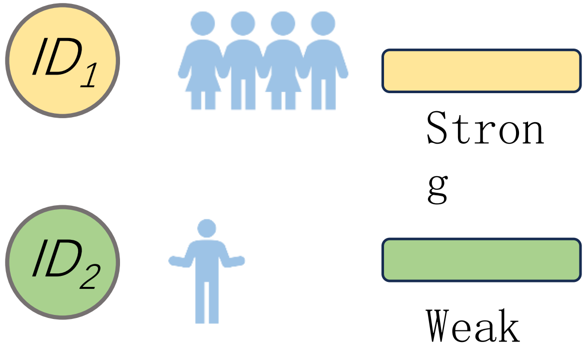
(1) Cold-start setting



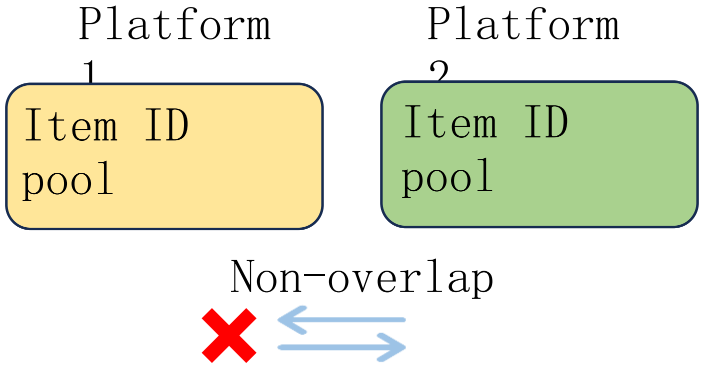
(2) Non-transferable

Background and Motivation

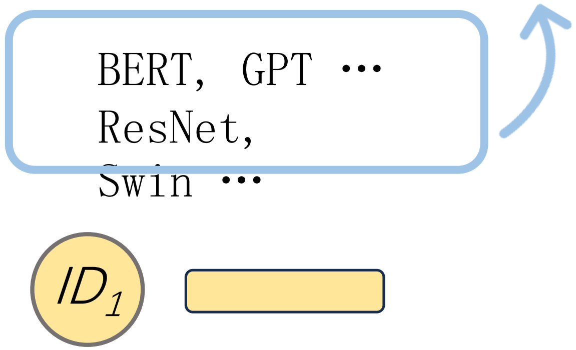
Key weaknesses of ID-based Recsys (IDRec)



(1) Cold-start setting



(2) Non-transferable



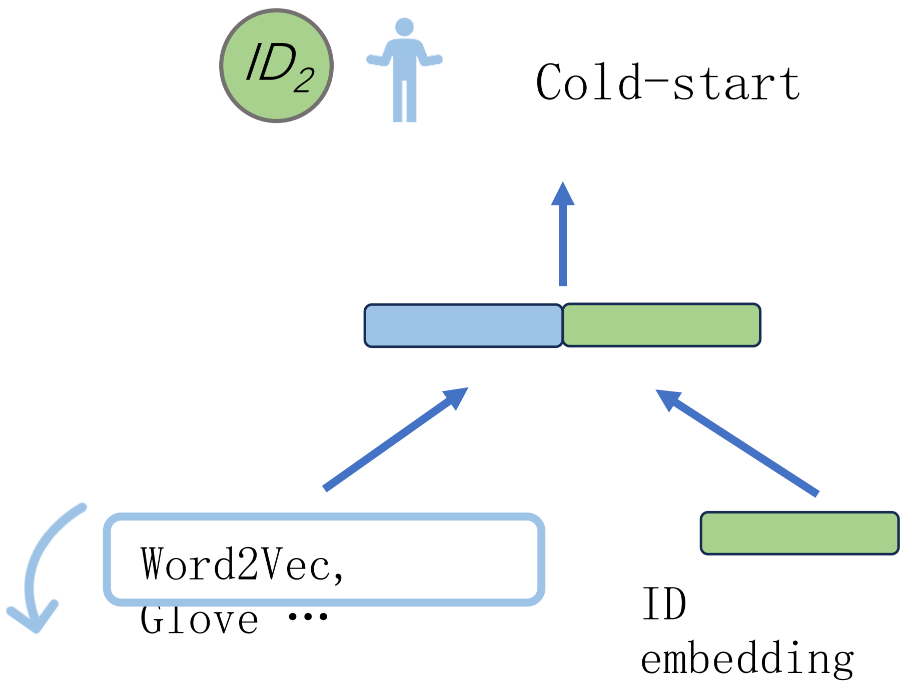
(3) Non-benefits

Background and Motivation

Modality-based Recsys (MoRec)

Background and Motivation

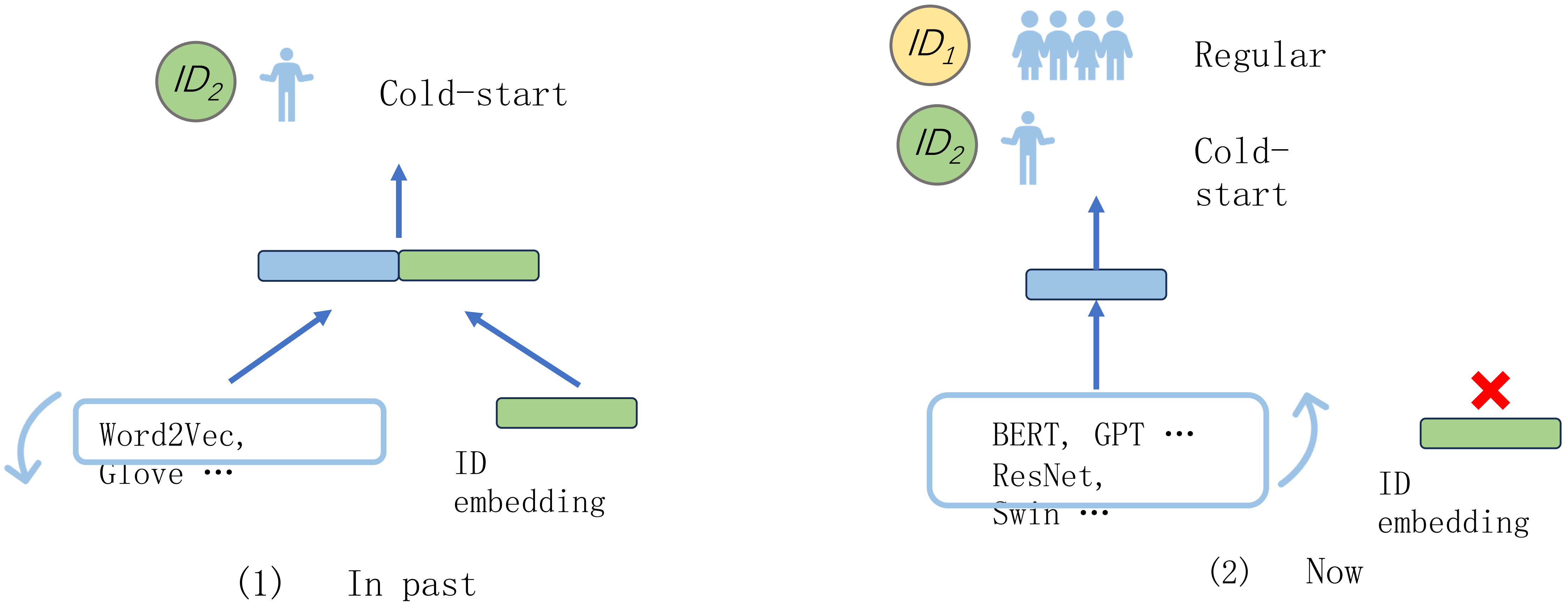
Modality-based Recsys (MoRec)



(1) In past

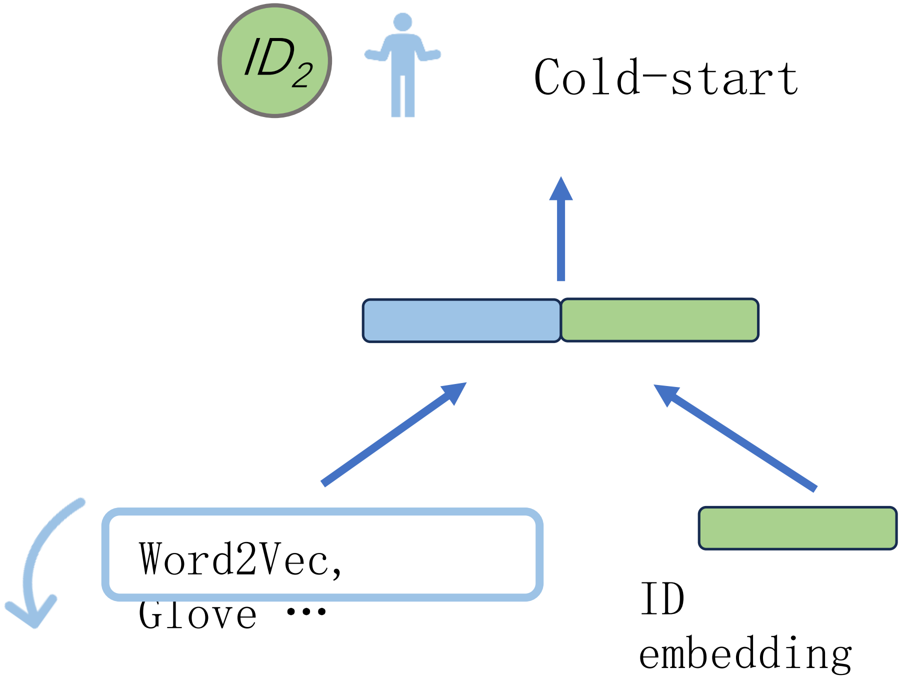
Background and Motivation

Modality-based Recsys (MoRec)

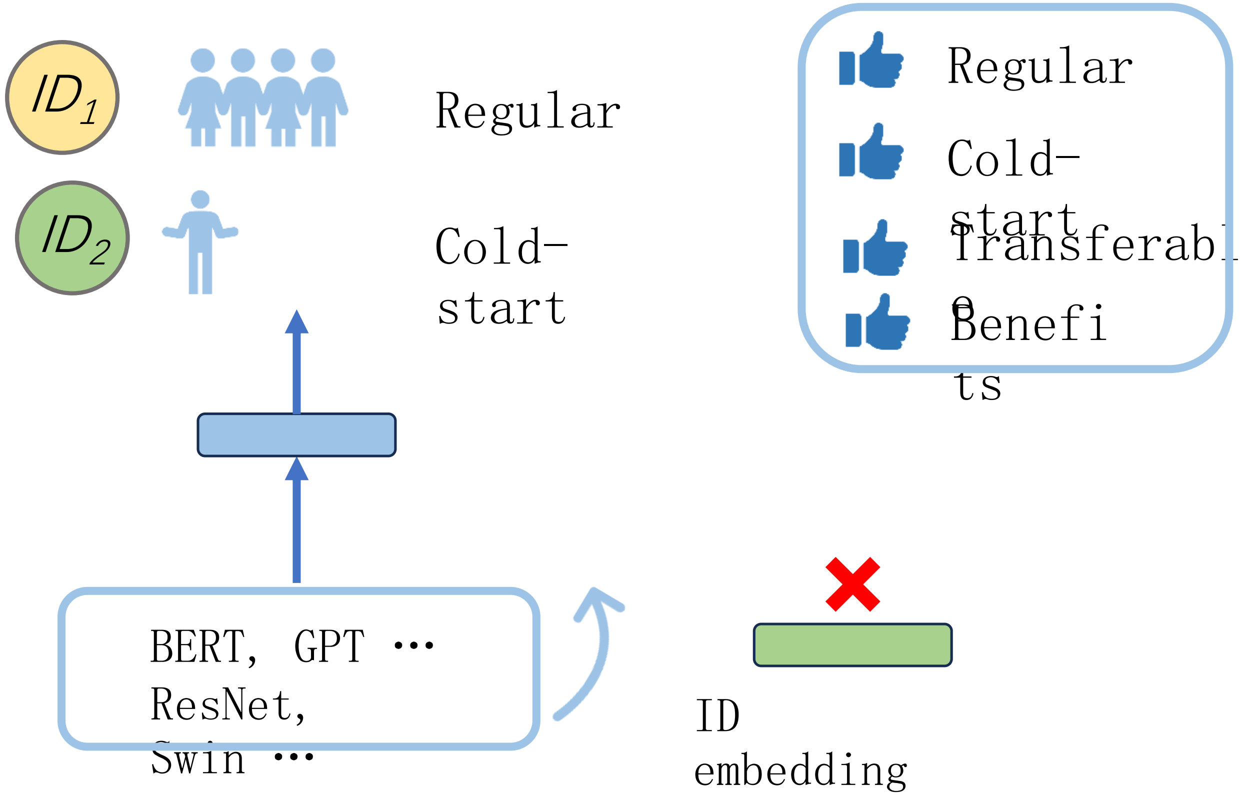


Background and Motivation

Modality-based Recsys (MoRec)



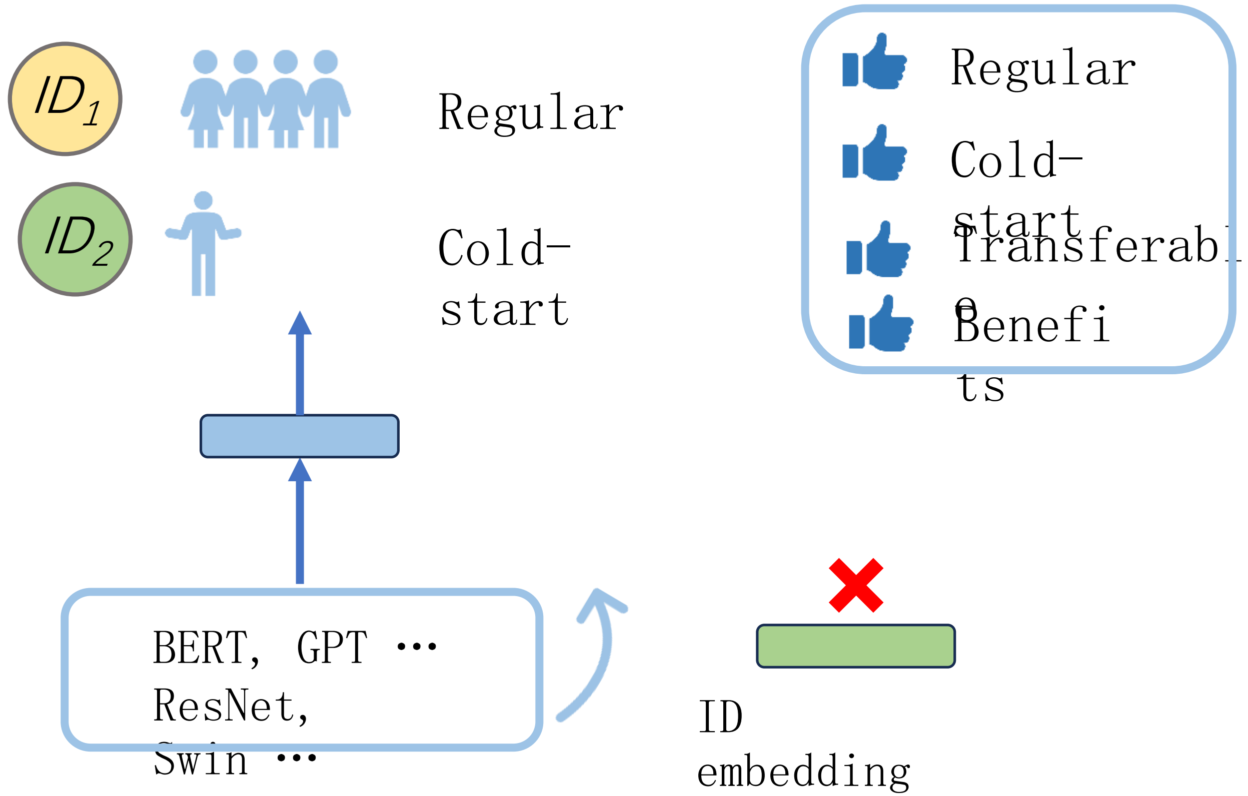
(1) In past



(2) Now

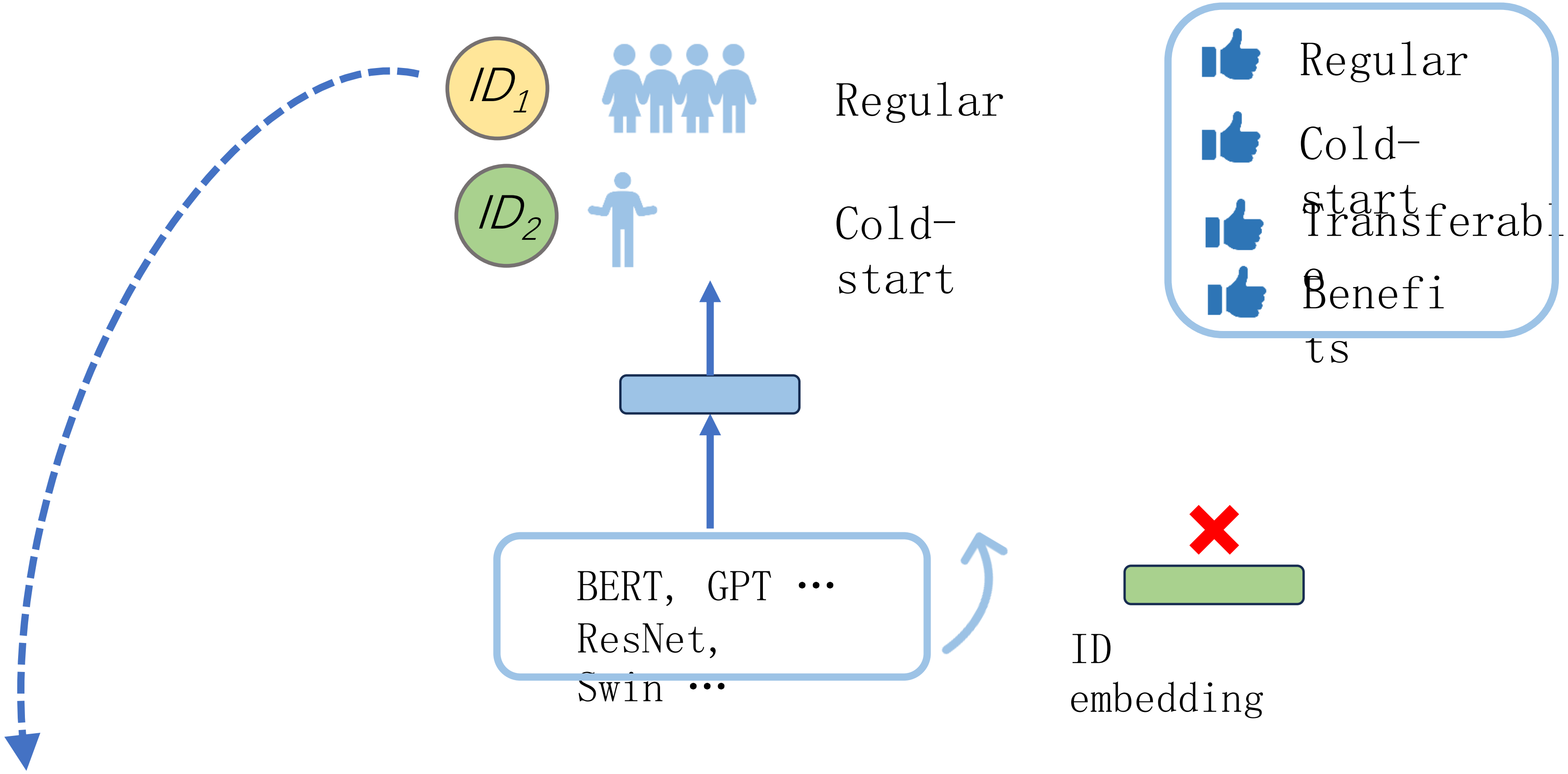
Background and Motivation

Rethink the potential of MoRec



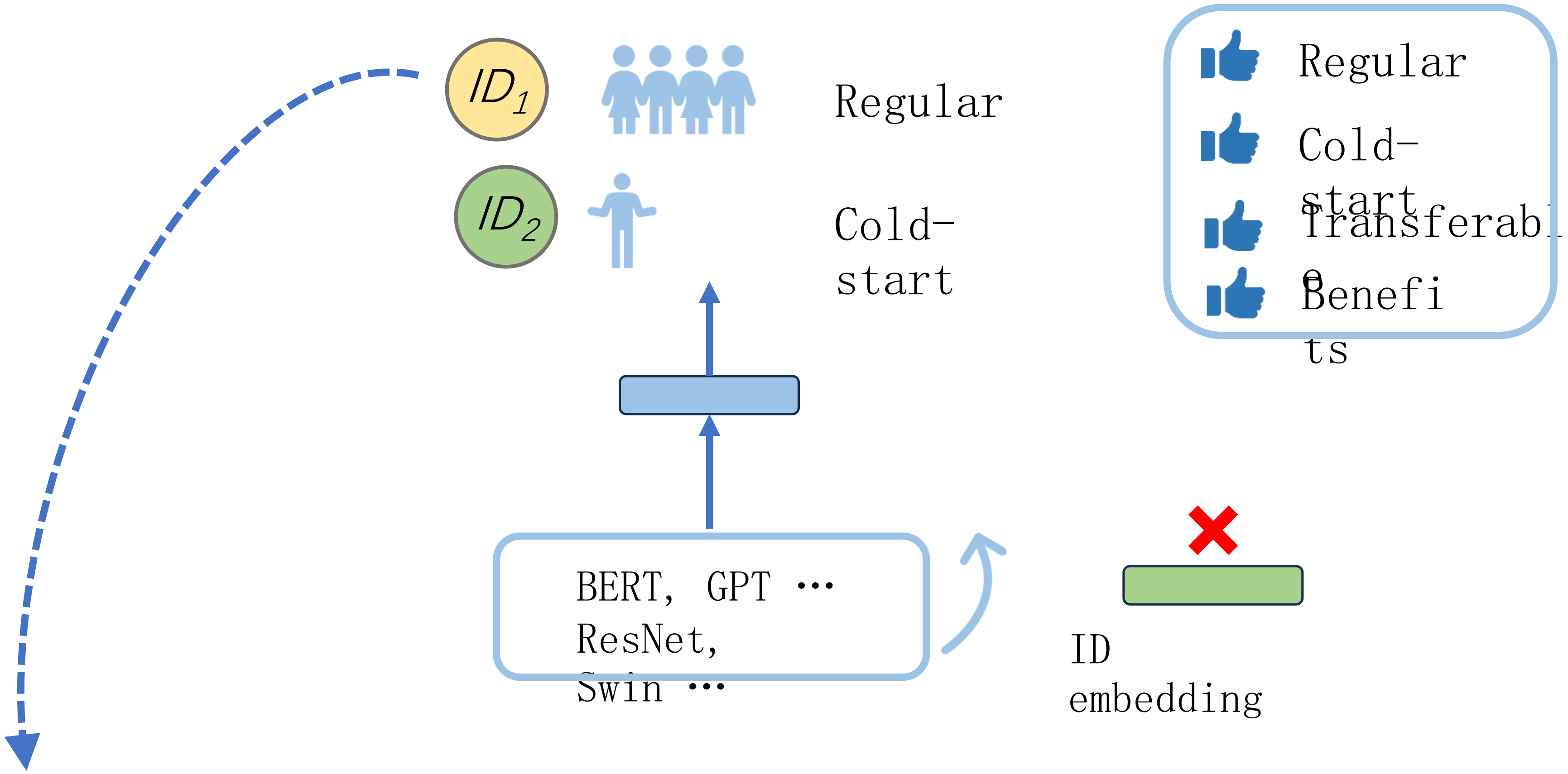
Background and Motivation

Rethink the potential of MoRec (1/4)



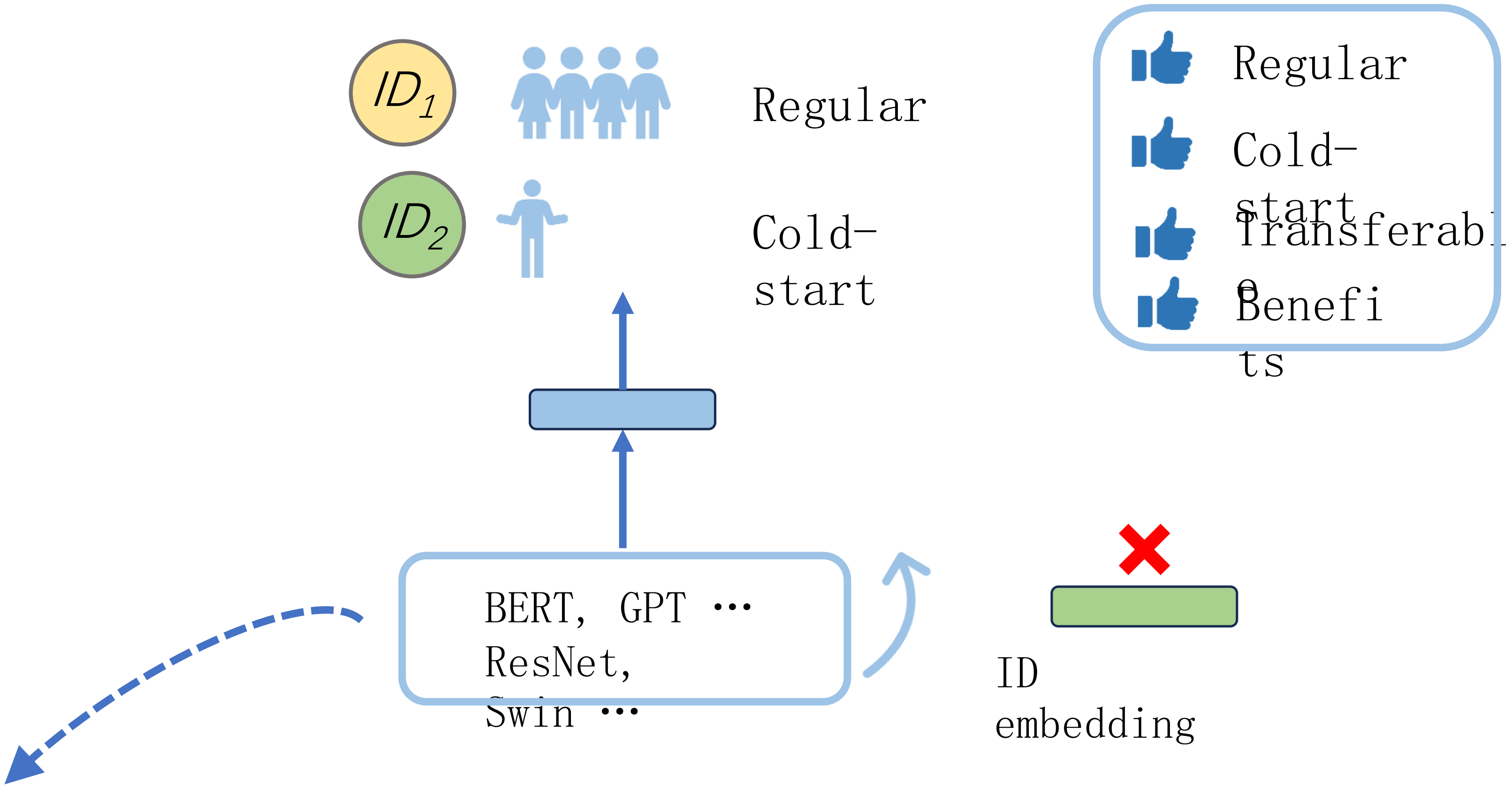
Background and Motivation

Rethink the potential of MoRec (1/4)

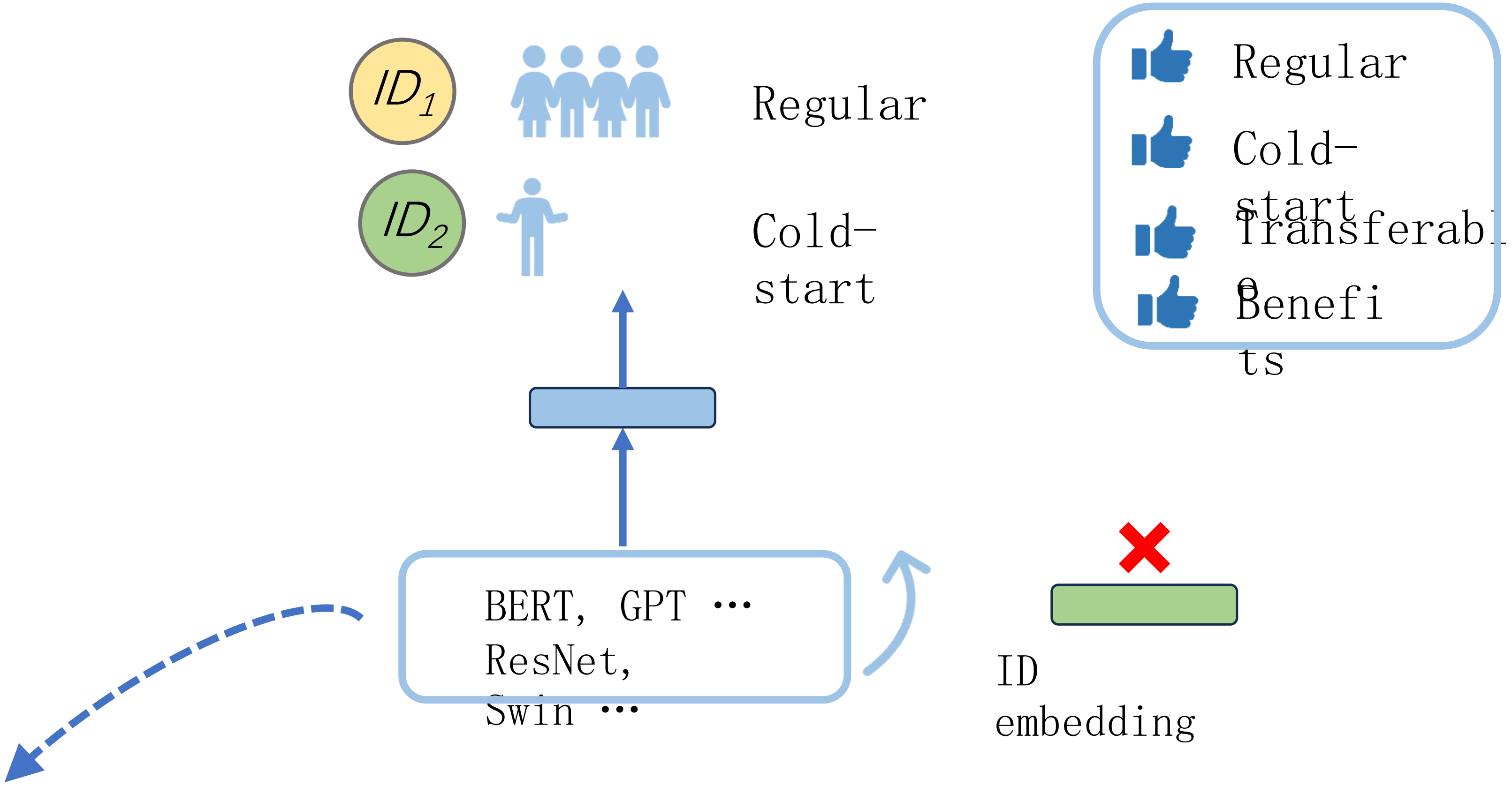


Q(i): Equipped with strong modality encoders, can MoRec be comparable to or even surpass IDRec in no-cold setting?

Rethink the potential of MoRec (2/4)



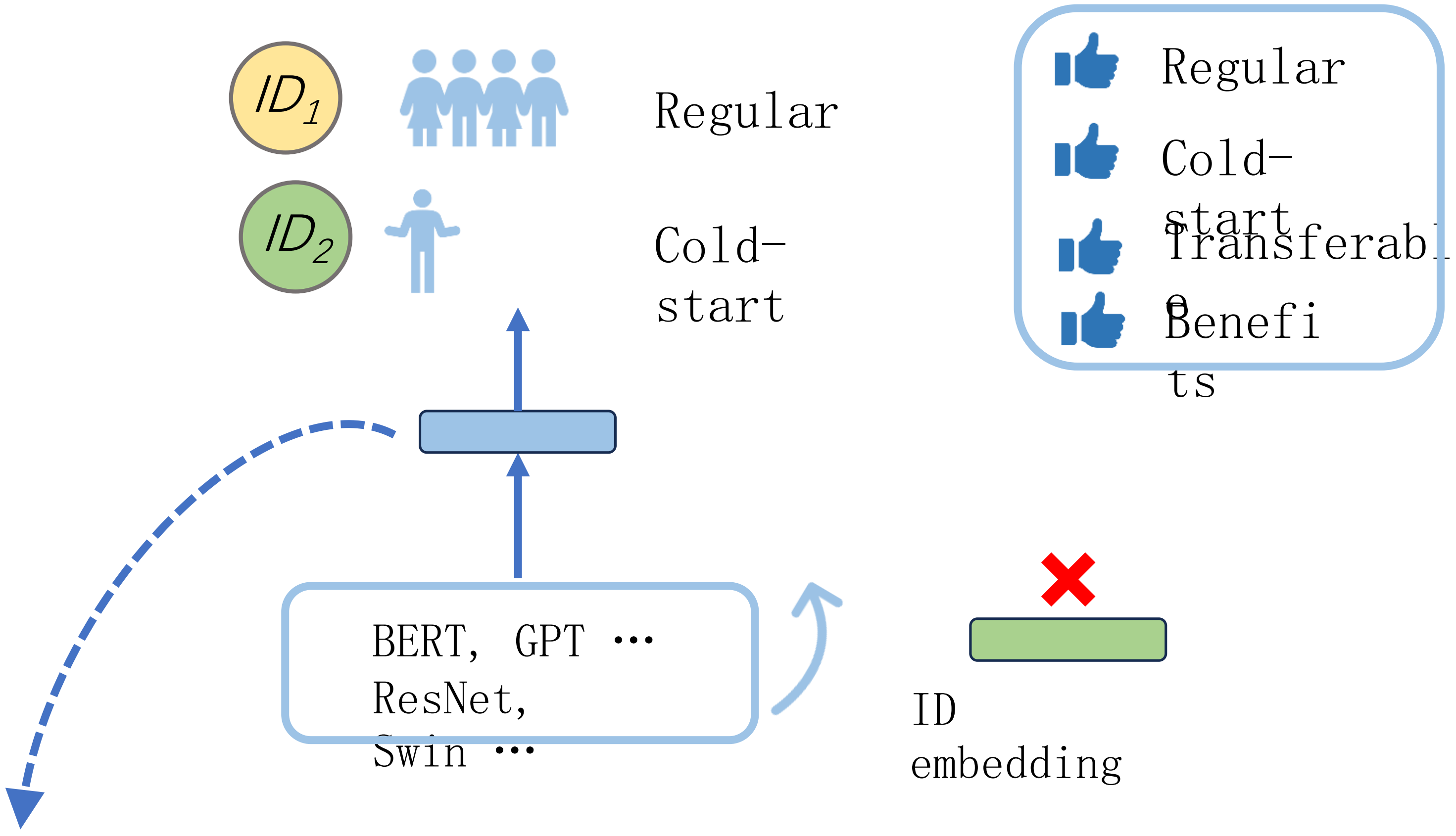
Rethink the potential of MoRec (2/4)



Q(ii): Can the recent technical advances developed in NLP and CV fields
translate into improvement in MoRec?

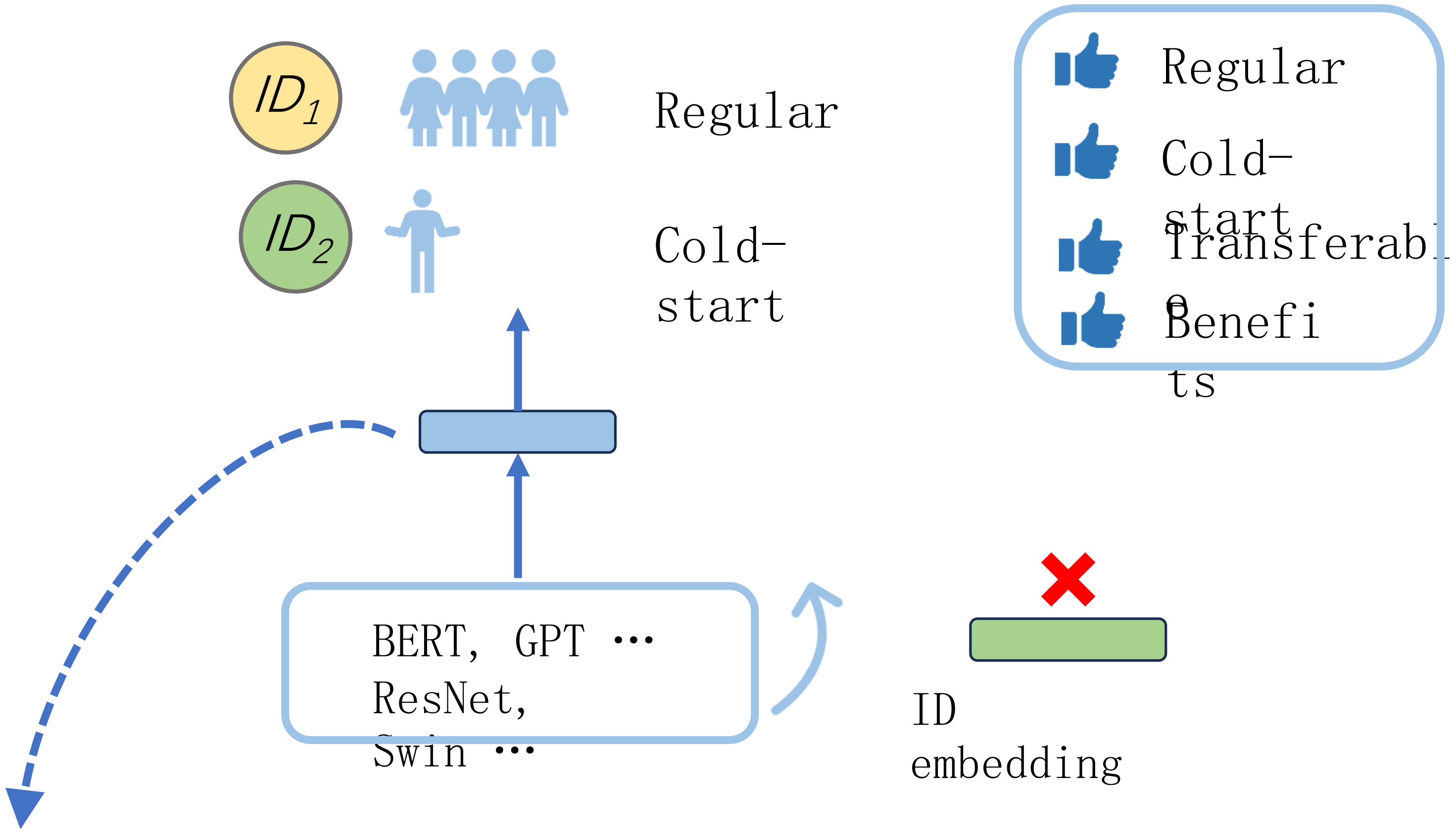
Background and Motivation

Rethink the potential of MoRec (3/4)



Background and Motivation

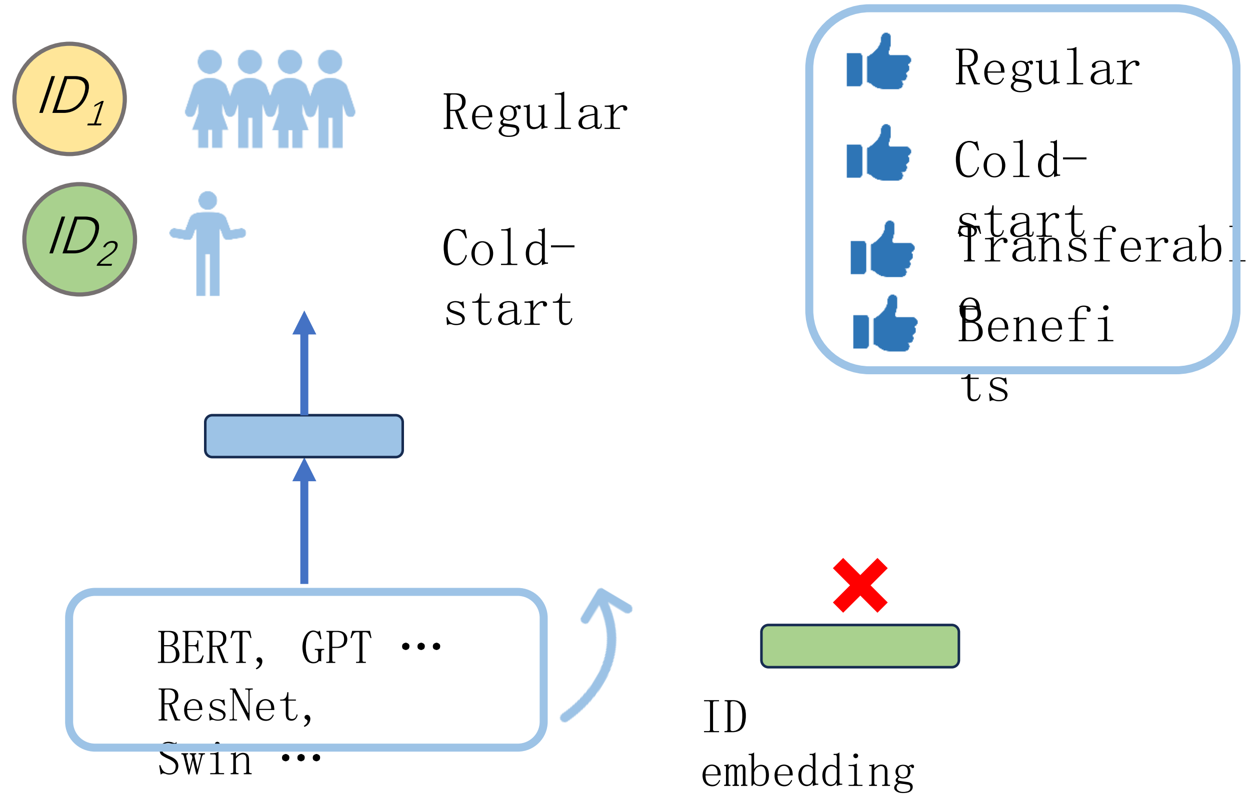
Rethink the potential of MoRec (3/4)



Q(iii): How can we effectively use the item modality representations derived from a pre-training NLP or CV encoder?

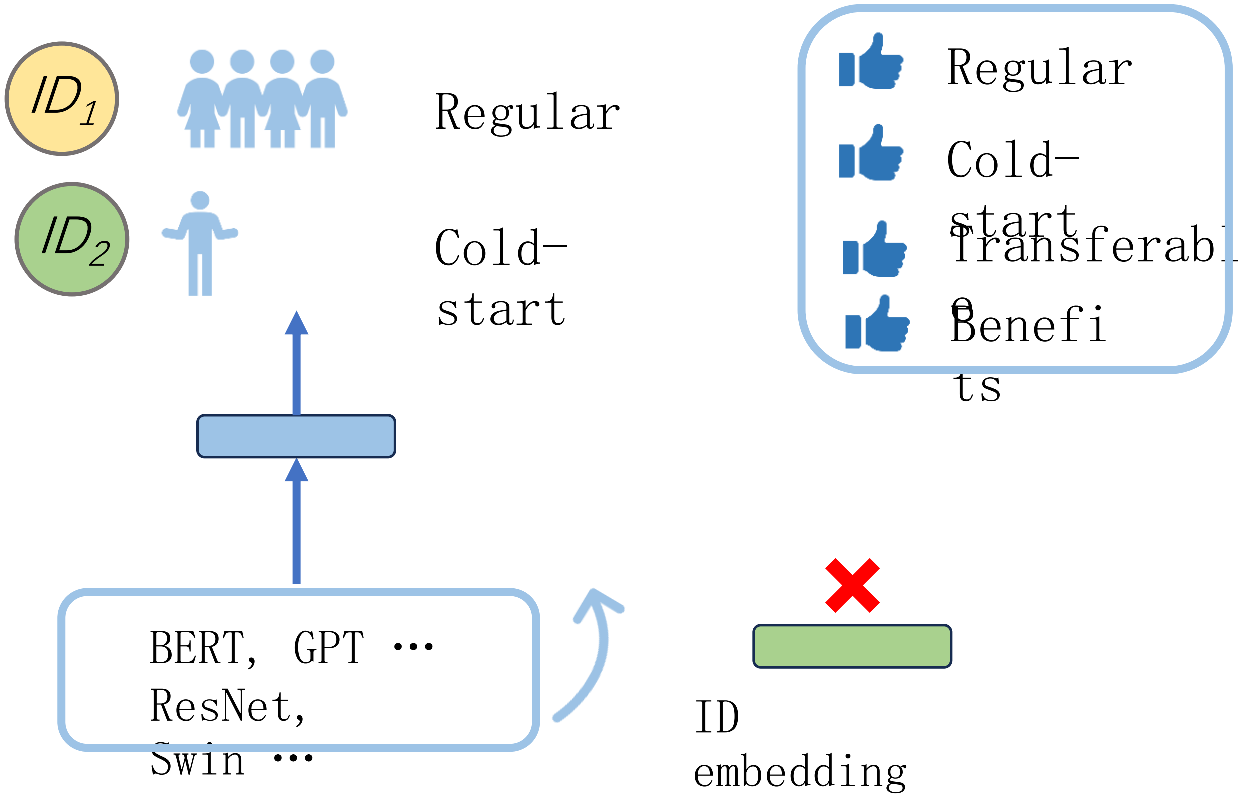
Background and Motivation

Rethink the potential of MoRec (4/4)



Background and Motivation

Rethink the potential of MoRec (4/4)



Q(iv): Several key challenges that remain unexplored for MoRec training.

Overview

Background and Motivation

Framework and Training Details

Experiments and Findings

Future Works

Framework

Framework

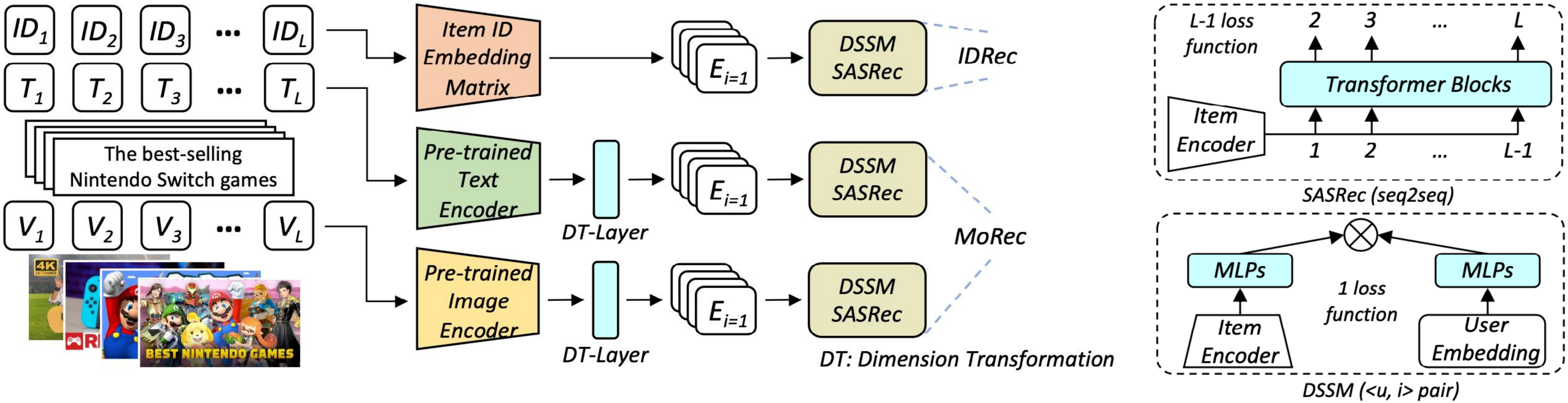
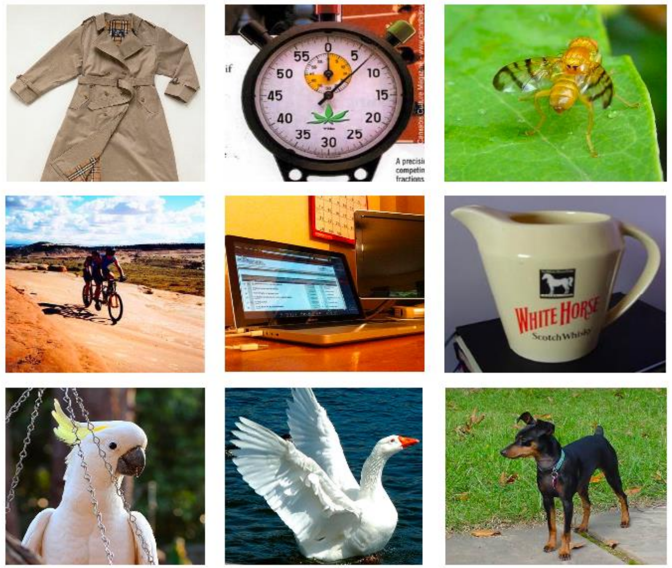


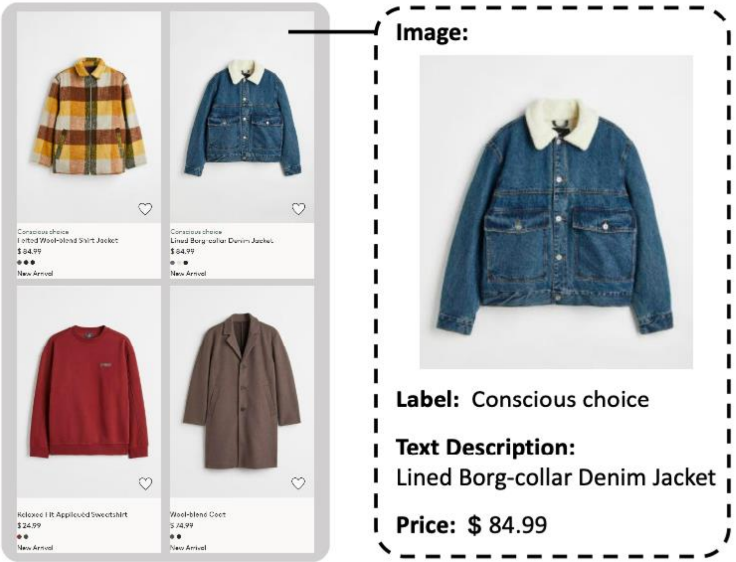
Illustration of IDRec vs MoRec.

Datasets and Loss Function

Datasets and Loss Function



(a) Item cases on ImageNet1K.

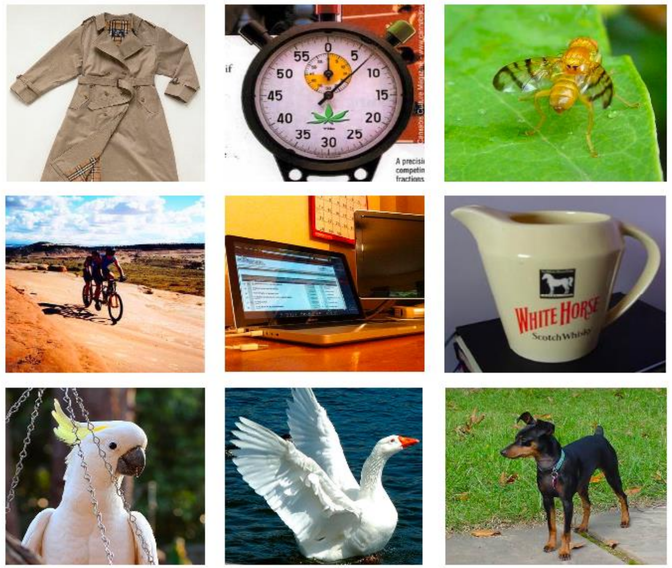


(b) Item cases on HM.

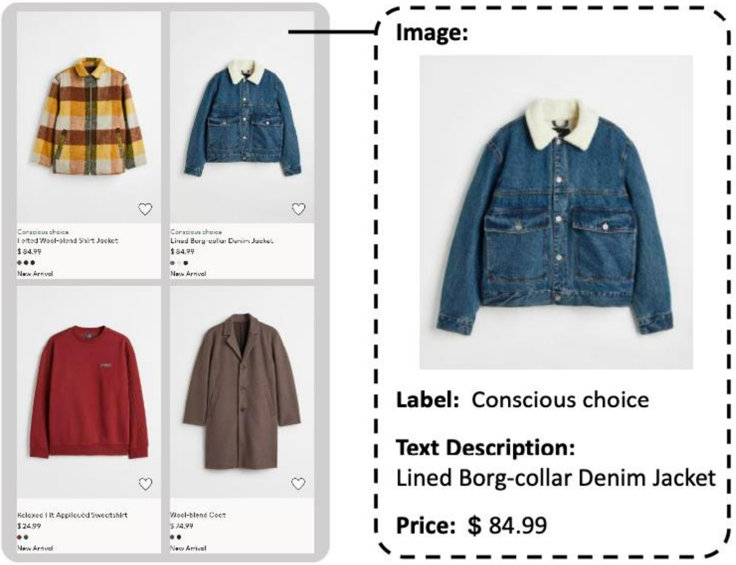
(c) Item cases on Bili.

Item cases on datasets.

Datasets and Loss Function



(a) Item cases on ImageNet1K.



(b) Item cases on HM.

(c) Item cases on Bili.

Item cases on datasets.

Binary cross entropy loss

$$\begin{cases} min - \sum_{u \in \mathcal{U}} \sum_{i \in [2,\ldots,L]} \{\log(\sigma(\hat{y}_{ui})) + \log(1 - \sigma(\hat{y}_{uj}))\} & \text{SASRec} \\ min - \sum_{<u,i,j> \in \mathcal{R}} \{\log(\sigma(\hat{y}_{ui})) + \log(1 - \sigma(\hat{y}_{uj}))\} & \text{DSSM} \end{cases}$$

Overview

Background and Motivation

Framework and Training Details

Experiments and Findings

Future Works

Q(i):

- Q(i): Equipped with strong modality encoders, can MoRec be comparable to or even surpass IDRec in no-cold setting?

Q(i):

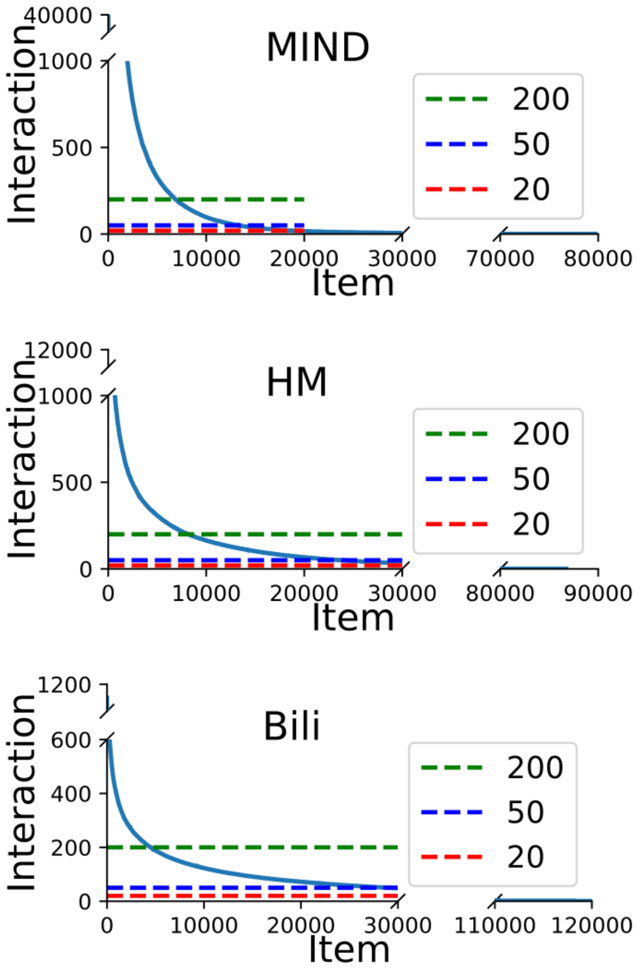
- Q(i): Equipped with strong modality encoders, can MoRec be comparable to or even surpass IDRec in no-cold setting?

Dataset	Metrics	DSSM			SASRec				Improv.
		IDRec	$\text{BERT}_{\text{base}}$	$\text{RoBERTa}_{\text{base}}$	IDRec	$\text{BERT}_{\text{small}}$	$\text{BERT}_{\text{base}}$	$\text{RoBERTa}_{\text{base}}$	
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)

Q(i):

- Q(i): Equipped with strong modality encoders, can MoRec be comparable to or even surpass IDRec in no-cold setting?



Dataset	MIND		HM		Bili	
	IDRec	$\text{BERT}_{\text{base}}$	IDRec	Swin-T	IDRec	Swin-T
Warm-20	20.12	**20.19**	7.89	**8.05**	3.48	**3.57**
Warm-50	20.65	**20.89**	**8.88**	8.83	**4.04**	4.02
Warm-200	**22.00**	21.73	**11.15**	11.10	**10.04**	9.98

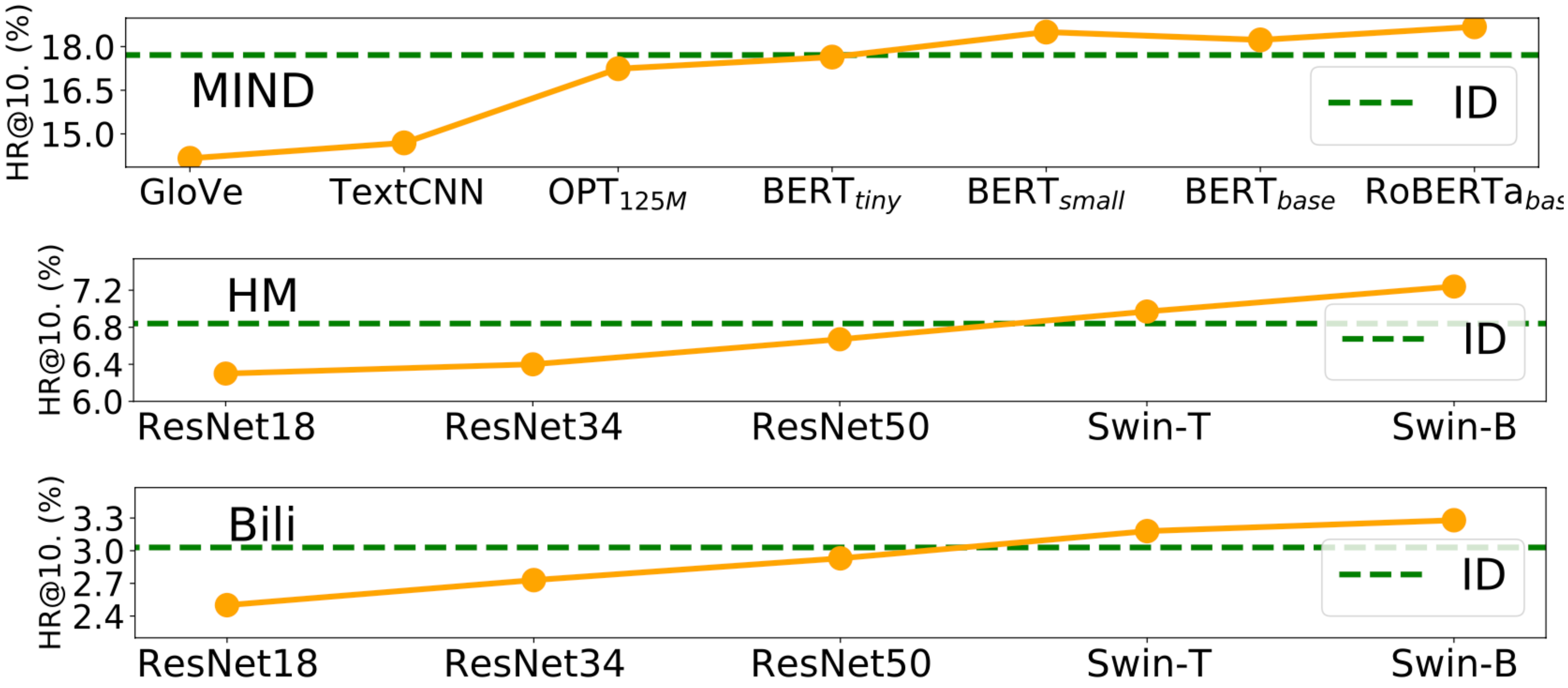
MoRec vs IDRec (Warm Item Settings)

Q(ii):

- Q(ii): Can the recent technical advances developed in NLP and CV fields translate into accuracy improvement in MoRec?

Q(ii):

- Q(ii): Can the recent technical advances developed in NLP and CV fields translate into accuracy improvement in MoRec?



Accuracy with different pre-trained ME in MoRec

Q(iii):

- Q(iii): How can we effectively use the item modality representations derived from an pre-training NLP or CV encoder?

Q(iii):

- Q(iii): How can we effectively use the item modality representations derived from an pre-training NLP or CV encoder?

Dataset	IDRec	ME	TS	TS-DNN					E2E
				2	6	8	10	12	
MIND	17.71	$BERT_{base}$	13.93	15.20	16.26	<u>16.66</u>	16.32	16.14	**18.23**
HM	6.84	ResNet50	4.03	4.64	<u>5.40</u>	5.39	<u>5.40</u>	5.02	6.67
		Swin-T	3.45	4.46	5.28	<u>5.55</u>	5.40	5.38	**6.97**
Bili	3.03	ResNet50	0.72	1.23	<u>1.62</u>	1.47	1.28	1.24	2.93
		Swin-T	0.79	1.40	1.81	<u>2.10</u>	1.95	1.64	**3.18**

E2E vs TS with additional MLP layers

Q(iv):

- Q(iv): Several key challenges that remain unexplored for MoRec training.

Q(iv):

- Q(iv): Several key challenges that remain unexplored for MoRec training.

Dataset	Method	#Param.	FLOPs	Time/E	MU	GPU
MIND	IDRec	47M	0.12G	2.7m	3G	V100-32G(1)
	$\text{BERT}_{\text{tiny}}$	11M	0.63G	10m	4G	V100-32G(1)
	$\text{BERT}_{\text{small}}$	35M	16G	42m	13G	V100-32G(1)
	$\text{BERT}_{\text{base}}$	116M	107G	102m	52G	V100-32G**(2)**
HM	IDRec	114M	1G	4.3m	5G	V100-32G(1)
	ResNet18	18M	40G	95m	23G	V100-32G(1)
	ResNet34	29M	81G	136m	30G	V100-32G(1)
	ResNet50	31M	91G	83m	80G	V100-32G**(4)**
	Swin-T	34M	96G	107m	157G	A100-40G**(4)**
	Swin-B	94M	333G	102m	308G	A100-40G**(8)**

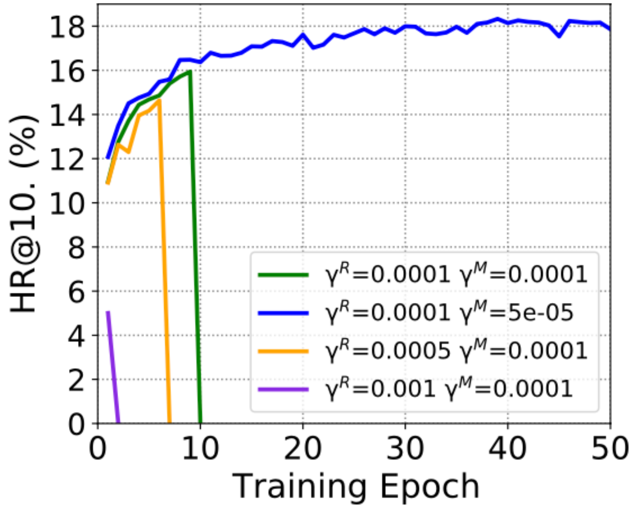
The training cost of IDRec and End2end MoRec

Q(iv):

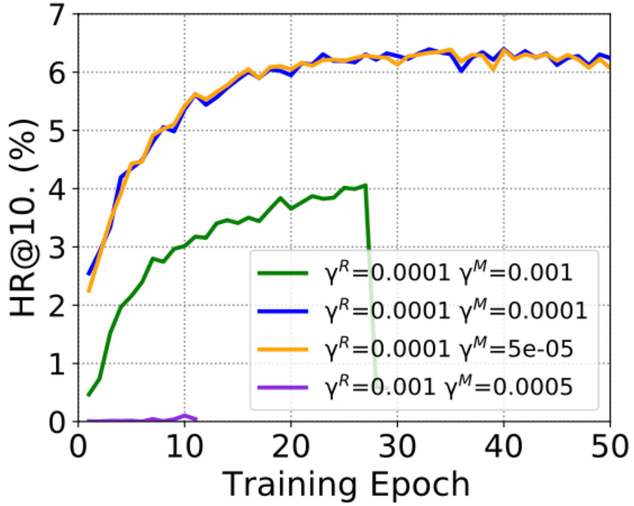
- Q(iv): Several key challenges that remain unexplored for MoRec training

Dataset	ME	TS			E2E		
		w/o	w/	Improv.	w/o	w/	Improv.
MIND	$\text{BERT}_{\text{base}}$	13.93	**14.68**	+5.38%	18.23	**18.63**	+2.19%
HM	MAE_{base}	2.50	**2.79**	+11.60%	7.03	**7.07**	+0.57%
Bili	MAE_{base}	0.57	0.57	0.00%	**3.18**	3.17	-0.31%

Performing a second round of pre-training for ME



(a) MoRec with $\text{BERT}_{\text{base}}$ on MIND.



(b) MoRec with Swin-T on HM.

Q(iv):

- Q(iv): Several key challenges that remain unexplored for MoRec training

Dataset	ME	IDRec	TS	ID+TS w/o ADD	ID+TS w/o CON	ID+TS w/ ADD	ID+TS w/ CON	Improv.	TS-DNN	ID+TS-DNN w/o ADD	ID+TS-DNN w/o CON	ID+TS-DNN w/ ADD	ID+TS-DNN w/ CON	Improv.	E2E	ID+E2E w/o ADD	ID+E2E w/o CON	Improv.
MIND	$\text{BERT}_{\text{base}}$	17.71	13.93	16.10	17.20	17.66	17.57	-0.28%	16.66	14.93	16.58	17.29	17.55	-0.90%	**18.23**	16.25	17.12	-6.09%
HM	Swin-T	6.84	3.45	5.75	4.89	5.37	5.40	-15.94%	5.55	5.27	4.00	4.77	5.11	-22.95%	**6.97**	5.40	4.95	-22.53%
Bili	Swin-T	3.03	0.79	3.01	2.61	3.02	2.86	-0.33%	2.10	2.86	2.35	2.50	2.72	-5.61%	**3.18**	2.94	2.55	-7.55%

co-training ID and modality.

Overview

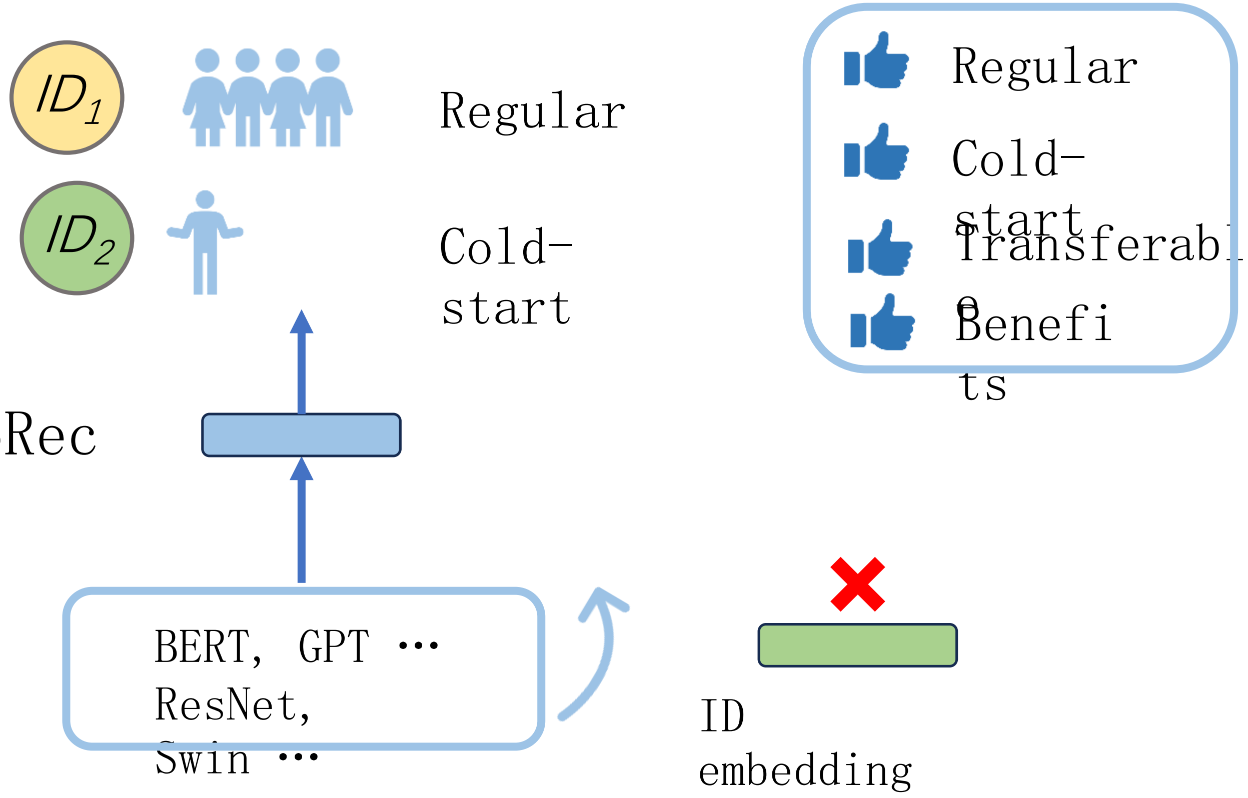
Background and Motivation

Framework and Training Details

Experiments and Findings

Future Works

Future Works



2. Multimodal MoRec

1. Text, vision, voice and video

3. Improved networks;
4. Scale up training data
5. Large Language Models: ChatGPT, GPT4

Where to Go Next for Recommender Systems? ID- vs. Modality-based Recommender Models Revisited

Zheng Yuan[1], Fajie Yuan[1], Yu Song[1], Youhua Li[1], Junchen Fu[1],
Fei Yang[2], Yunzhu Pan[1], Yongxin Ni[1]

Q&A

西湖大學
WESTLAKE UNIVERSITY