Learning to Warm Up Cold Item Embeddings for Cold-start Recommendation with Meta Scaling and Shifting Networks

Yongchun Zhu[1,2,3] , Ruobing Xie[3], Fuzhen Zhuang[4,5], Kaikai Ge[3], Ying Sun[1,2],

Xu Zhang[3], Leyu Lin[3], Juan Cao[1,2]

1. Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS) , Institute of Computing Technology, CAS, Beijing, China

2. University of Chinese Academy of Sciences, Beijing, China

3. WeChat Search Application Department, Tencent, China

4. Institute of Artificial Intelligence, Beihang University, Beijing 100191, China

5. Xiamen Data Intelligence Academy of ICT, CAS, China

Motivation

Item Cold-start Problem

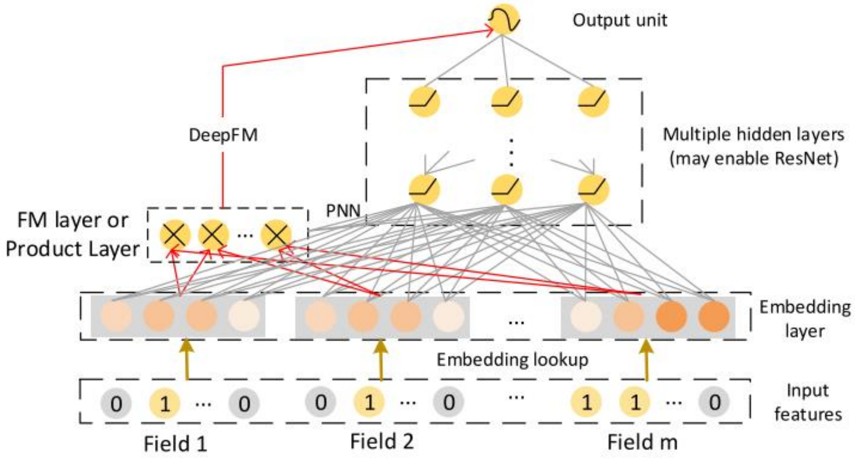
New item

How to recommend these new(cold-start) items to potential users?

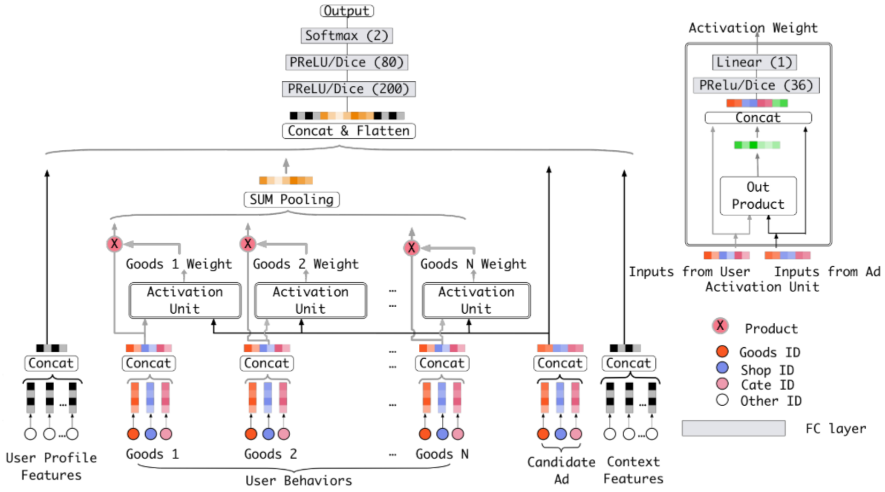
No History Data For Recommendation

Motivation

Embedding Technology



DeepFM



DIN

Deep recommendation models:

- Embedding layer
- Deep model

Item ID embedding is transformed from an item identifier (item ID).

Motivation

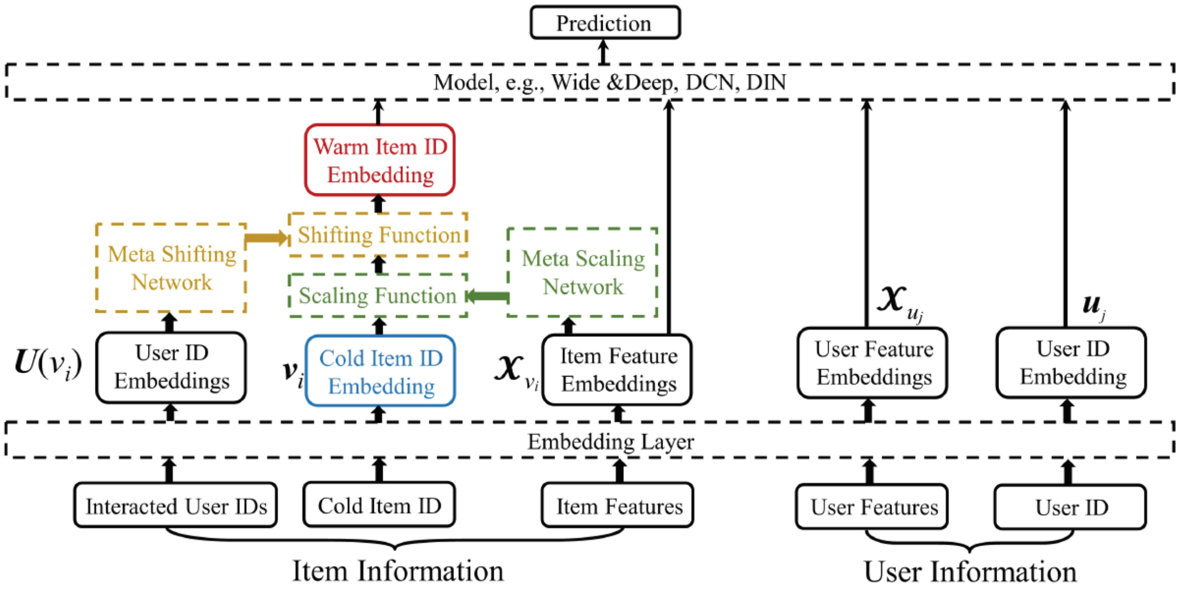
Challenges

- A gap is existing between the cold ID embedding and the deep model.
- The cold item ID embedding would be seriously influenced by noise.

Solution

- Fast adaptation (Meta Scaling Network)
- Alleviate the influence of noise (Meta Shifting Network)

Meta Warm Up Framework (MWUF)



- Common initialization
- Meta Scaling Network

$$\tau_{v_i}^{scale} = h(\mathcal{X}_{v_i}; w_{scale}),\ \tau^{scale} \in \mathbb{R}^k,$$

- Meta Shifting Network

$$\tau_{v_i}^{shift} = g(\mathcal{G}(U(v_i)); w_{shift}),\ \tau^{shift} \in \mathbb{R}^k,$$

Warm Embedding

$$v_i^{warm} = v_i \odot \tau_{v_i}^{scale} + \tau_{v_i}^{shift}$$

INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

WeChat

Meta Warm Up Framework (MWUF)

Overall procedure

- Train the deep model and embedding layer
- Train two meta networks with the common initial embeddings

$$\hat{y}^{cold} = f(\hat{v}_i, \mathcal{X}_{v_i}, u_j, \mathcal{X}_{u_j}; \theta)$$

$$\hat{y}^{warm} = f(\hat{v}_i^{warm}, \mathcal{X}_{v_i}, u_j, \mathcal{X}_{u_j}; \theta)$$

$$\min_{\phi_{id}^{new}} \mathcal{L}^{cold}, \qquad \min_{w^{scale}, w^{shift}} \mathcal{L}^{warm}$$

Algorithm 1 Meta Warm Up Framework (MWUF).

Input: f_θ: A pre-trained model.
Input: $h_{w_{scale}}$, $g_{w_{shift}}$: Two meta networks.
Input: ϕ_{id}^{new}: An initialized item ID embedding layer.
Input: $\mathcal{D}$: A dataset sorted by timestamp.

(1) randomly initialize $h_{w_{scale}}$, $g_{w_{shift}}$.
(2) **while** not converge **do**:
(3) Sample batch of samples $\mathcal{B}$ from $\mathcal{D}$
(4) Obtain cold ID embeddings v of items in $\mathcal{B}$
(5) Obtain warm ID embeddings v^{warm} by Equation(5)
(6) Evaluate $\mathcal{L}^{cold}$ with v
(7) Update ϕ_{id}^{new} by minimizing $\mathcal{L}^{cold}$
(8) Evaluate $\mathcal{L}^{warm}$ with v^{warm}
(9) Update $h_{w_{scale}}$, $g_{w_{shift}}$ by minimizing $\mathcal{L}^{warm}$
(10) **end while**

Experiments

Datasets

- MovieLens-1M
- Taobao Display Ad Click
- CIKM2019 EComm AI

Metrics: AUC, RelaImpr

Baselines

- Popular CF methods: FM, Wide & Deep, PNN, DCN, AFN
- Cold-start methods: DropoutNet, MeLU, MetaEmb
- Sequential methods: GRU4Rec, DIN

Experiments

Research questions

- RQ1 How do different methods (popular CF models, sequential methods, cold-start methods) perform in the cold-start setting?
- RQ2 How does MWUF upon various deep recommendation models perform?
- RQ3 What are the effects of Initialization, Meta Scaling, and Meta Shifting Networks in our proposed MWUF?
- RQ4 Can the initialization methods (MetaEmb, MWUF) alleviate the cold-start problem?

Dataset splits

- Old items
- New items

Implementation Details

- Pre-train the base model with the data of old items.
- Train extra modules or initialize the new item ID.
- Compute evaluation metrics on the warm-a set; (**cold phase**)
- Update the embeddings of new items IDs with warm-a set and compute evaluation metrics on the warm-b set; (**warm-a phase**)
- Update the embeddings of new items IDs with warm-b set and compute evaluation metrics on the warm-c set; (**warm-b phase**)
- Update the embeddings of new items IDs with warm-c set and compute evaluation metrics on the test set; (**warm-c phase**)

Experiments

Results (RQ1)

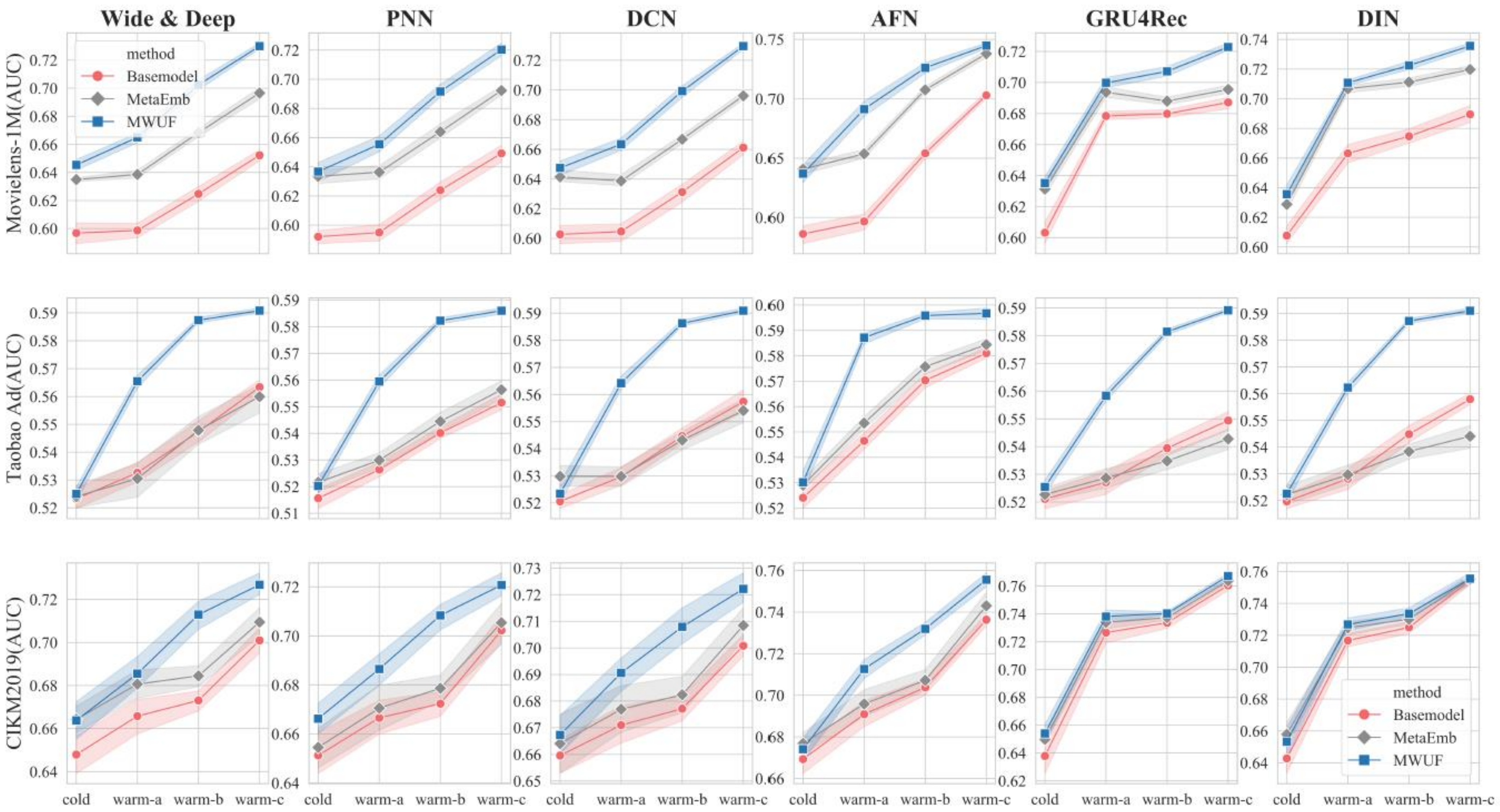
- The effectiveness of SOTA deep CF method.
- The effectiveness of various kinds of cold-start approaches.
- The effectiveness of sequential models.
- Recommendation performance on different phases.
- The effectiveness of MWUF.

Dataset	Methods	cold AUC	cold RelaImpr	warm-a AUC	warm-a RelaImpr	warm-b AUC	warm-b RelaImpr	warm-c AUC	warm-c RelaImpr
MovieLens-1M	FM	0.5250	-74.2%	0.5296	-70.0%	0.5406	-67.4%	0.5525	-65.5%
	Wide & Deep	0.5968	0.0%	0.5987	0.0%	0.6247	0.0%	0.6523	0.0%
	PNN	0.5920	-5.0%	0.5949	-3.9%	0.6240	-0.6%	0.6493	-2.0%
	DCN	0.6027	6.1%	0.6046	6.0%	0.6312	5.2%	0.6612	5.8%
	AFN	0.5862	-11.0%	0.5966	-2.1%	0.6541	23.6%	0.7029	33.2%
	DropoutNet	0.6432	47.9%	0.6499	51.9%	0.6565	25.5%	0.6628	6.9%
	MeLU	**0.6434**	**48.1%**	0.6400	41.8%	0.6643	31.8%	0.6760	15.6%
	MetaEmb	0.6369	41.4%	0.6349	36.7%	0.6686	35.2%	0.6983	30.2%
	GRU4Rec	0.6032	6.6%	0.6784	80.7%	0.6799	44.3%	0.6872	22.9%
	DIN	0.6077	11.3%	0.6632	65.3%	0.6747	40.1%	0.6895	24.4%
	MWUF(Wide & Deep)	0.6339	38.3%	0.6569	59.0%	0.6999	60.3%	0.7273	49.2%
	MWUF(GRU4Rec)	0.6298	34.1%	**0.6962***	**98.8%**	0.7033	63.0%	0.7160	41.8%
	MWUF(AFN)	0.6370	41.5%	0.6913	93.8%	**0.7261***	**81.3%**	**0.7447***	**60.7%**
Taobao Display AD	FM	0.5020	-91.4%	0.5064	-80.4%	0.5096	-79.7%	0.5126	-80.1%
	Wide & Deep	0.5233	0.0%	0.5326	0.0%	0.5474	0.0%	0.5634	0.0%
	PNN	0.5156	-33.0%	0.5264	-19.0%	0.5401	-15.4%	0.5515	-18.8%
	DCN	0.5206	-11.6%	0.5296	-9.2%	0.5446	-5.9%	0.5574	-9.5%
	AFN	0.5241	3.4%	0.5465	42.6%	0.5703	48.3%	0.5810	27.8%
	DropoutNet	0.5214	-8.2%	0.5318	-2.5%	0.5515	8.6%	0.5655	3.3%
	MeLU	0.5226	-3.0%	0.5223	-31.6%	0.5256	-46.0%	0.5360	-43.2%
	MetaEmb	0.5250	7.3%	0.5345	5.8%	0.5506	6.8%	0.5633	-0.2%
	GRU4Rec	0.5211	-9.4%	0.5271	-16.9%	0.5394	-16.9%	0.5494	-22.1%
	DIN	0.5197	-15.5%	0.5281	-13.8%	0.5448	-5.5%	0.5580	-8.5%
	MWUF(Wide & Deep)	0.5255	9.4%	0.5718	120.2%	0.5890	87.8%	0.5919	45.0%
	MWUF(GRU4Rec)	0.5208	-10.7%	0.5626	92.0%	0.5842	77.6%	0.5867	36.8%
	MWUF(AFN)	**0.5302**	**29.6%**	**0.5872***	**167.5%**	**0.5958***	**102.1%**	**0.5965***	**45.3%**
CIKM2019	FM	0.5386	-73.9%	0.5420	-74.7%	0.5510	-70.5%	0.5598	-70.2%
	Wide & Deep	0.6479	0.0%	0.6657	0.0%	0.6730	0.0%	0.7010	0.0%
	PNN	0.6512	2.2%	0.6665	0.5%	0.6723	-0.4%	0.7022	0.6%
	DCN	0.6596	7.9%	0.6710	3.2%	0.6771	2.4%	0.7008	-0.1%
	AFN	0.6693	14.5%	0.6909	15.2%	0.7038	17.8%	0.7363	17.6%
	DropoutNet	0.6555	5.1%	0.6719	3.7%	0.6854	7.2%	0.6980	-1.5%
	MeLU	0.6595	7.8%	0.6697	2.4%	0.6762	1.8%	0.7032	1.1%
	MetaEmb	0.6642	11.0%	0.6746	5.4%	0.6795	3.8%	0.7095	4.2%
	GRU4Rec	0.6378	-6.8%	0.7266	36.8%	0.7337	35.1%	0.7635	31.1%
	DIN	0.6428	-3.4%	0.7168	30.8%	0.7250	30.1%	0.7545	26.6%
	MWUF(Wide & Deep)	0.6637	10.7%	0.6855	11.9%	0.7097	21.2%	0.7236	11.2%
	MWUF(GRU4Rec)	0.6540	4.1%	**0.7381***	**43.7%**	**0.7404**	**39.0%**	**0.7672**	**32.9%**
	MWUF(AFN)	**0.6741**	**17.7%**	0.7126	28.3%	0.7320	34.1%	0.7556	27.2%

Experiments

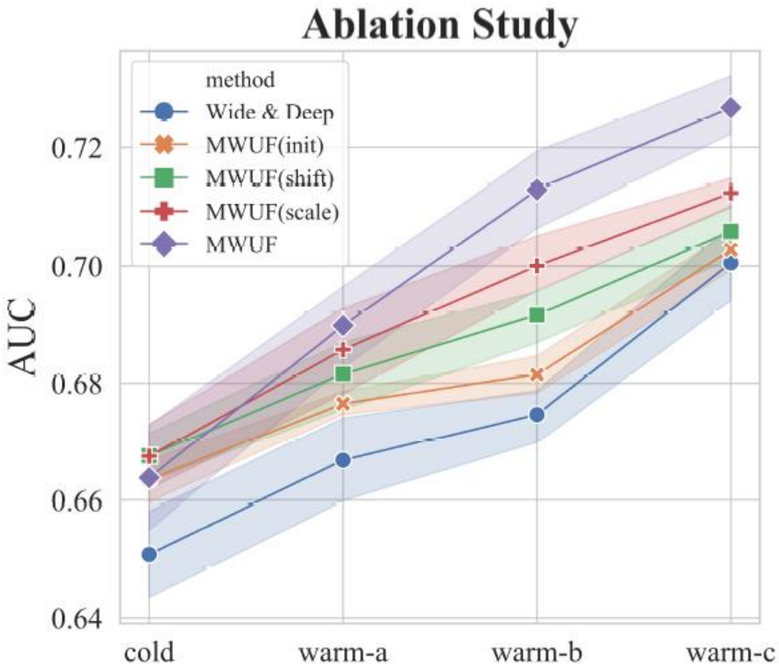
Generalization Experiments (RQ2)

- Compatibility.
- The effectiveness of MWUF.

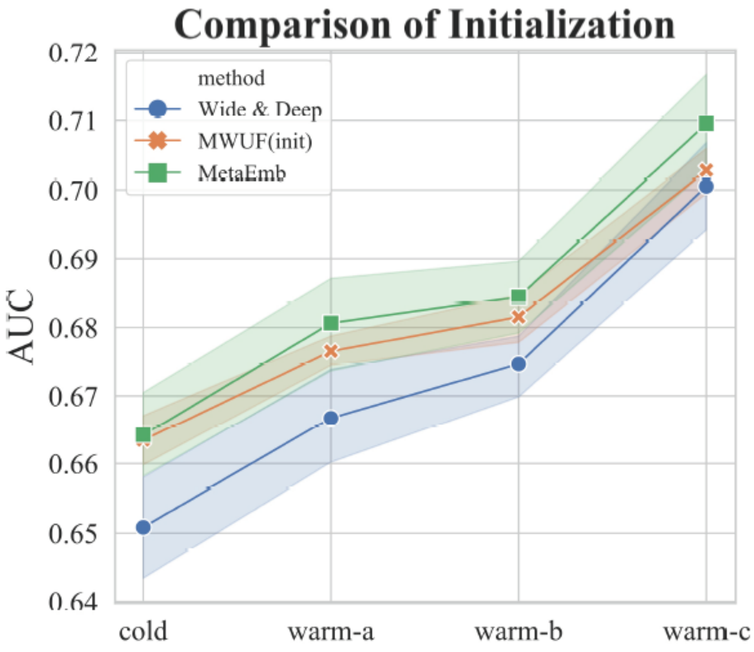


Experiments

Ablation Study (RQ3)



Comparison of Initialization (RQ4)



中国科学院计算技术研究所
INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES
WeChat

Thanks Q & A

Name: Yongchun Zhu | Email: zhuyongchun18s@ict.ac.cn

中国科学院计算技术研究所
INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

WeChat