Warm Up Cold-Start Advertisements

Improving CTR predictions via Learning to Learn ID embeddings

Feiyang Pan[1], Shuokai Li[1], Xiang Ao[1], Pingzhong Tang[2], Qing He[1]

[1] Institute of Computing Technology, CAS
[2] Tsinghua University

Feiyang Pan

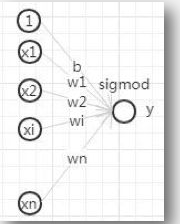
24 July 2019

What is CTR prediction？

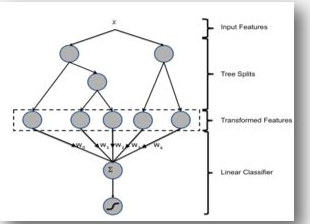
Binary Classification

Input : {ad, user, some contexts, … …}

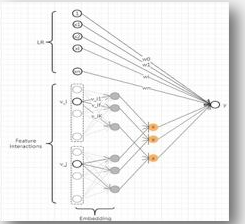
Label : {1, 0} → click or not



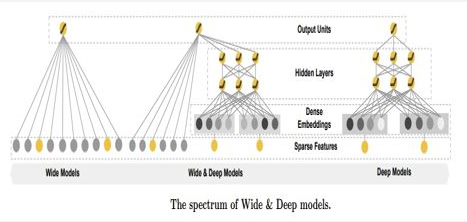
LR



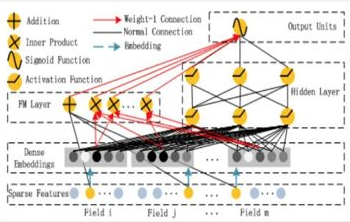
GBDT+LR



FM



Wide & Deep Learning

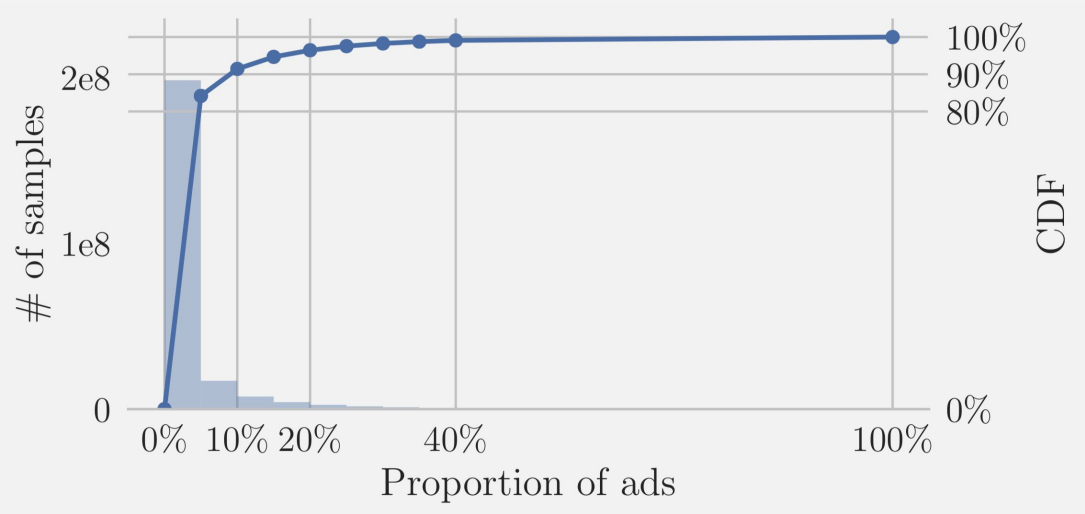


DeepFM

…

What is the cold-start problem？

The model is not familiar with **new / small** ads (or users).



KDD cup 2012 search ads dataset

5% of the ads accounted for nearly 90% of the samples

pCTR = f(embedding of the ad ID, ad features, contexts)

For new ads:

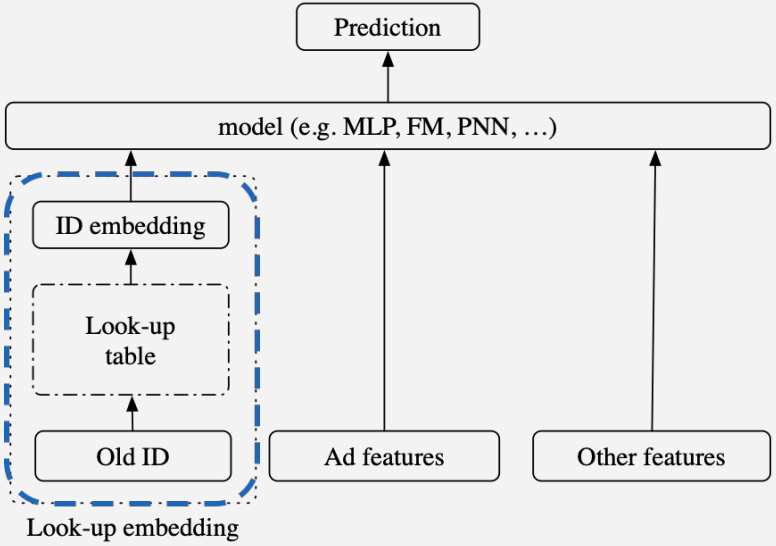
→ No labeled sample.

→ Unknown ID embedding.

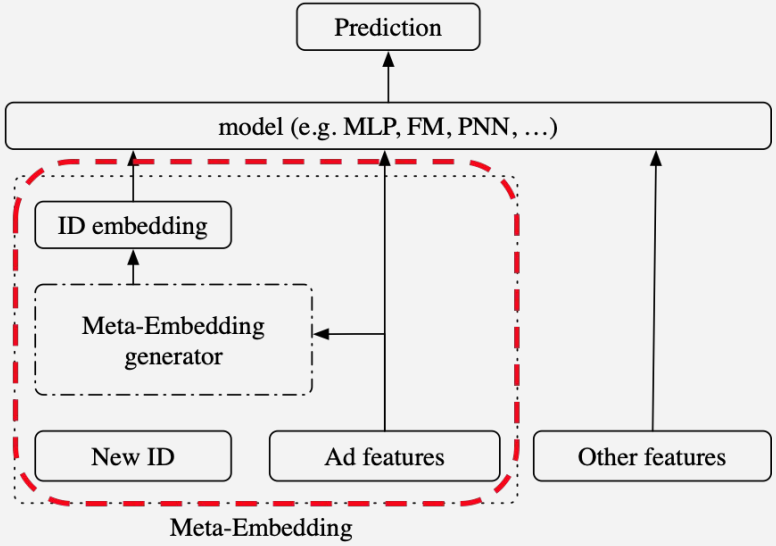
→ Inaccurate CTR prediction.

Meta-Embedding

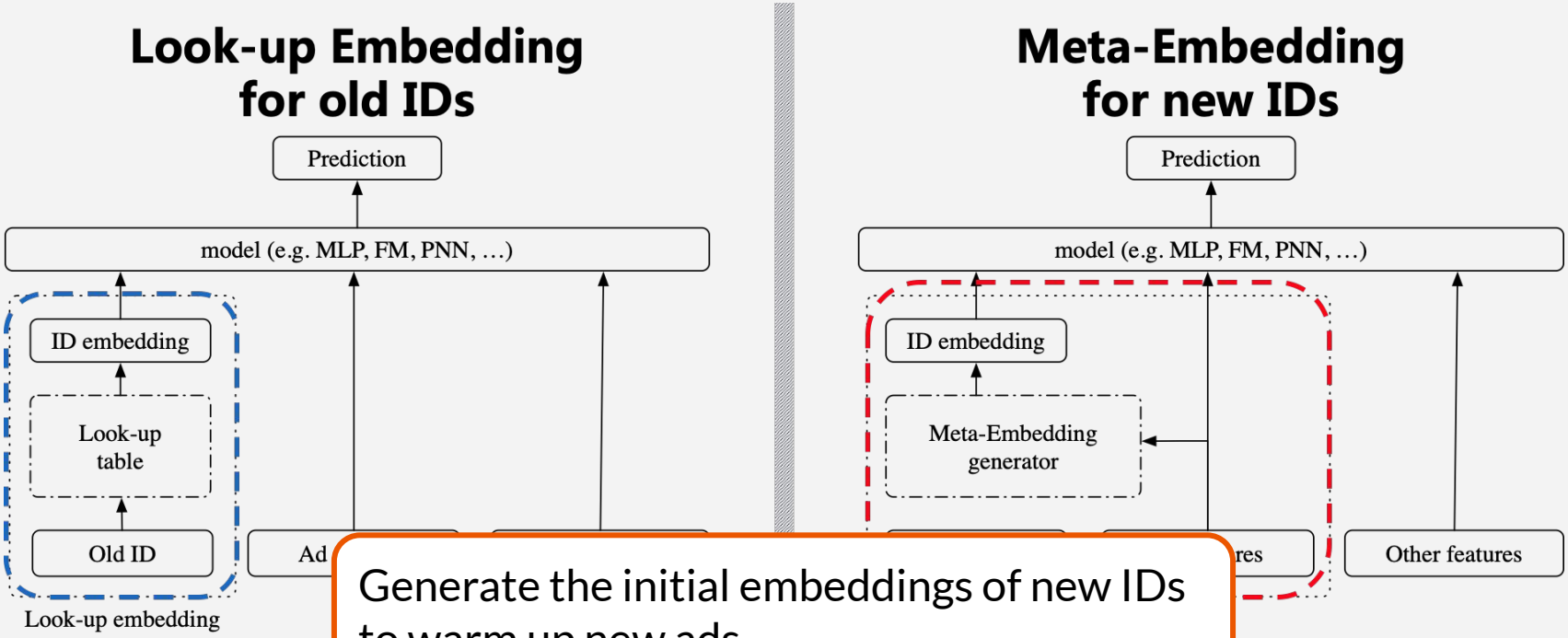
Look-up Embedding for old IDs



Meta-Embedding for new IDs

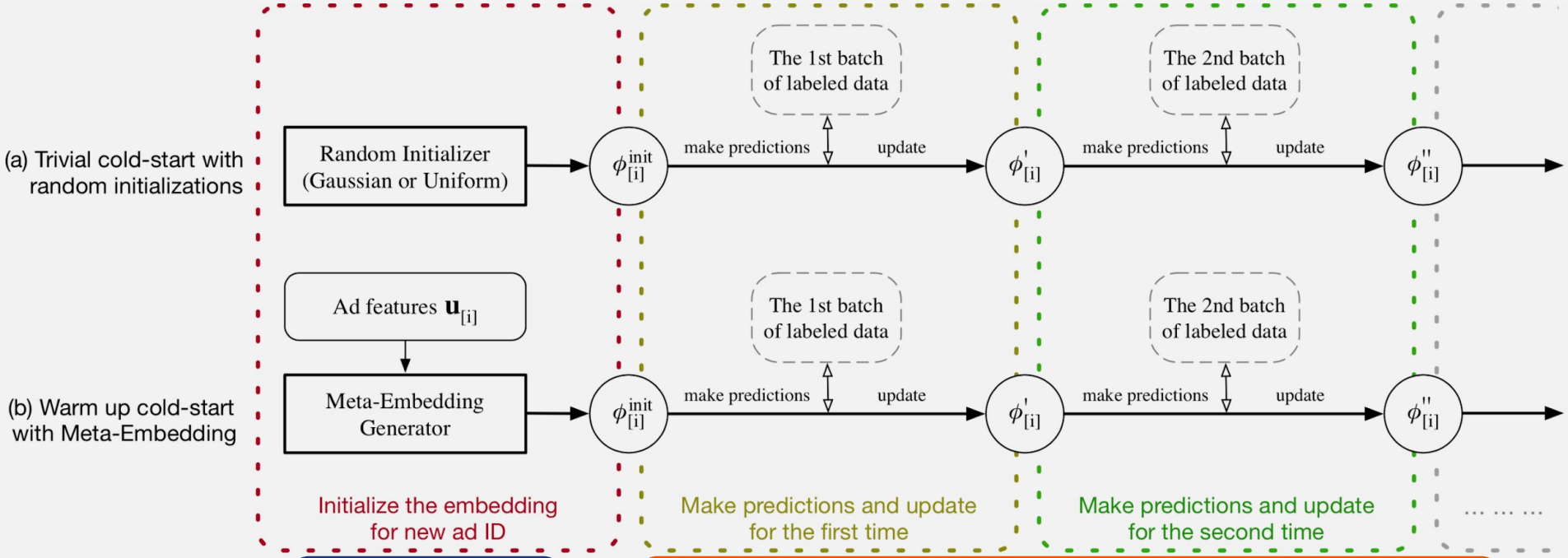


Meta-Embedding



Generate the initial embeddings of new IDs to warm up new ads.

How to use it?



Initialization (Offline)

Make predictions & update the embedding (Online)

Learning

Two phases & Two goals (for new ads) :

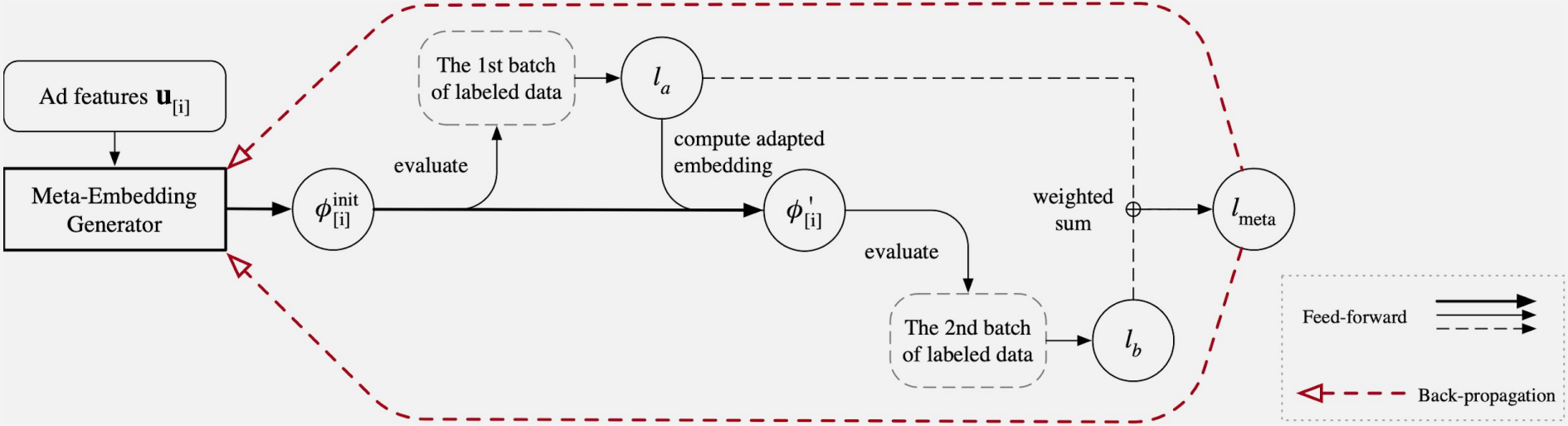
(a) **cold-start phase:**

give good predictions for new ads without labeled data.

(b) **warm-up phase:**

learn quickly with a small number of labeled examples.

Learning



$$loss_{meta} = \alpha\, loss_a + (1-\alpha)\, loss_b$$

End-to-end training.

Learning

Algorithm 1 Train Meta-Embedding by SGD

Input: f_θ: the pre-trained base model.
Input: $\mathcal{I}$: the set of all existing IDs.
Input: α: hyper-parameter, the coefficient for meta-loss.
Input: a, b: step sizes.

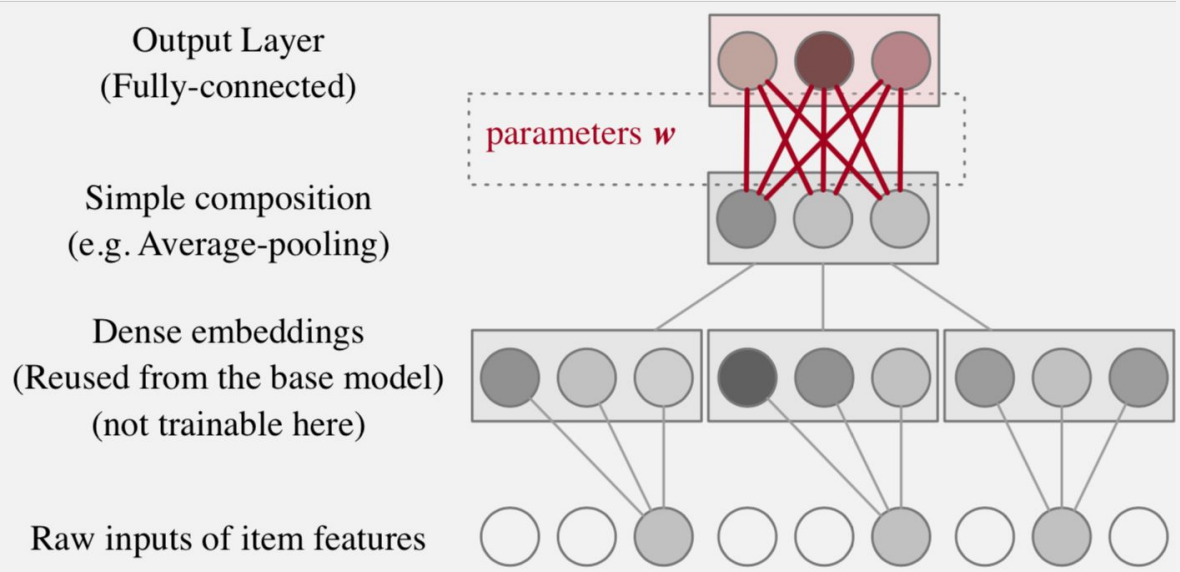
Randomly initialize $\boldsymbol{w}$
while not done **do**
  Randomly sample n IDs $\{\mathbf{i}_1, \ldots, \mathbf{i}_n\}$ from $\mathcal{I}$
  for $\mathbf{i} \in \{\mathbf{i}_1, \ldots, \mathbf{i}_n\}$ **do**
    Generate the initial embedding: $\phi^{\text{init}}_{[\mathbf{i}]} = h_{\boldsymbol{w}}(\mathbf{u}_{[\mathbf{i}]})$
    Sample mini-batches $\mathcal{D}^a_{[\mathbf{i}]}$ and $\mathcal{D}^b_{[\mathbf{i}]}$ each with K samples
    Evaluate loss $l_a(\phi^{\text{init}}_{[\mathbf{i}]})$ on $\mathcal{D}^a_{[\mathbf{i}]}$
    Compute adapted embedding: $\phi'_{[\mathbf{i}]} = \phi^{\text{init}}_{[\mathbf{i}]} - a \frac{\partial l_a(\phi^{\text{init}}_{[\mathbf{i}]})}{\partial \phi^{\text{init}}_{[\mathbf{i}]}}$
    Evaluate loss $l_b(\phi'_{[\mathbf{i}]})$ on $\mathcal{D}^b_{[\mathbf{i}]}$
    Compute loss: $l_{\text{meta},\mathbf{i}} = \alpha l_a(\phi^{\text{init}}_{[\mathbf{i}]}) + (1-\alpha) l_b(\phi'_{[\mathbf{i}]})$
  Update $\boldsymbol{w} \leftarrow \boldsymbol{w} - b \sum_{\mathbf{i} \in \{\mathbf{i}_1, \ldots, \mathbf{i}_n\}} \frac{\partial l_{\text{meta},\mathbf{i}}}{\partial \boldsymbol{w}}$

End-to-end training with SGD.

Can be applied in both the offline and online settings.

Details

The basic structure of the embedding generator:



Experiments set-up

For each new ads, we split 3 mini-batches for simulating cold-start, others are held-out for testing.

Table 1: The statistics for data splitting

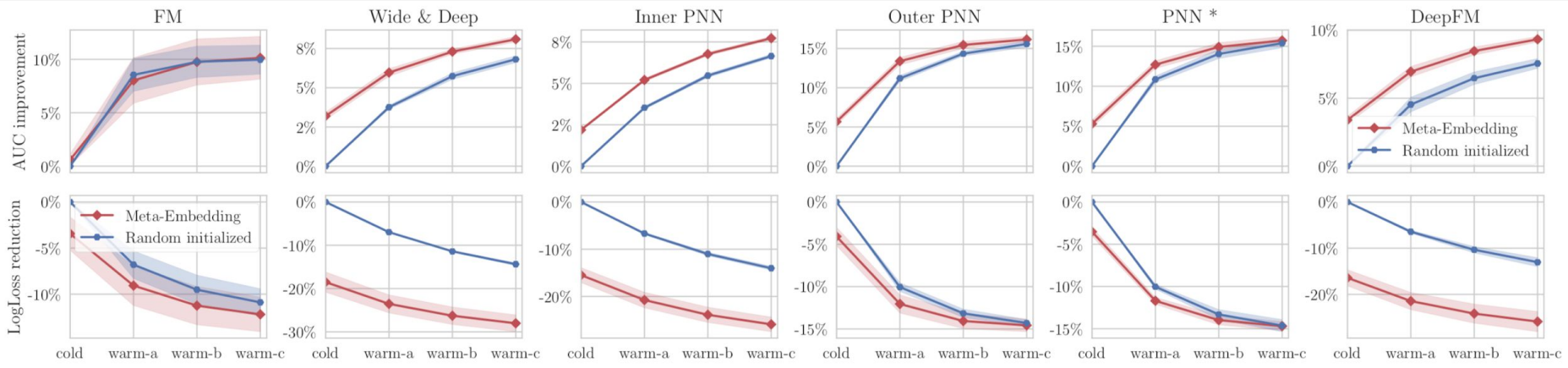
Dataset	Minibatch size K	Old ads			New ads		
		# of IDs	# of samples	# of samples used to train Meta-Embedding	# of IDs	# of samples	# of samples in the hold-out set
MovieLens-1M	20	1127	0.76 M	0.09 M	1058	0.19 M	0.12 M
Tencent CVR data	200	572	49.33 M	0.45 M	443	5.00 M	4.74 M
KDD Cup 2012	200	6534	148.55 M	5.22 M	9299	28.71 M	23.13 M

Experiment pipeline :

1. Pre-train the base model with the data of old ads ;
2. Train the Meta-Embedding with the training data ;
3. Generate initial embeddings of new ad IDs with (random initialization or Meta-Embedding) :
4. Update the embeddings with batch-a and compute evaluation metrics on the hold-out set ;
5. Update the embeddings with batch-b and compute evaluation metrics on the hold-out set ;
6. Update the embeddings with batch-c and compute evaluation metrics on the hold-out set ;

Evaluation metrics: Improvements on Log-loss and the AUC score.

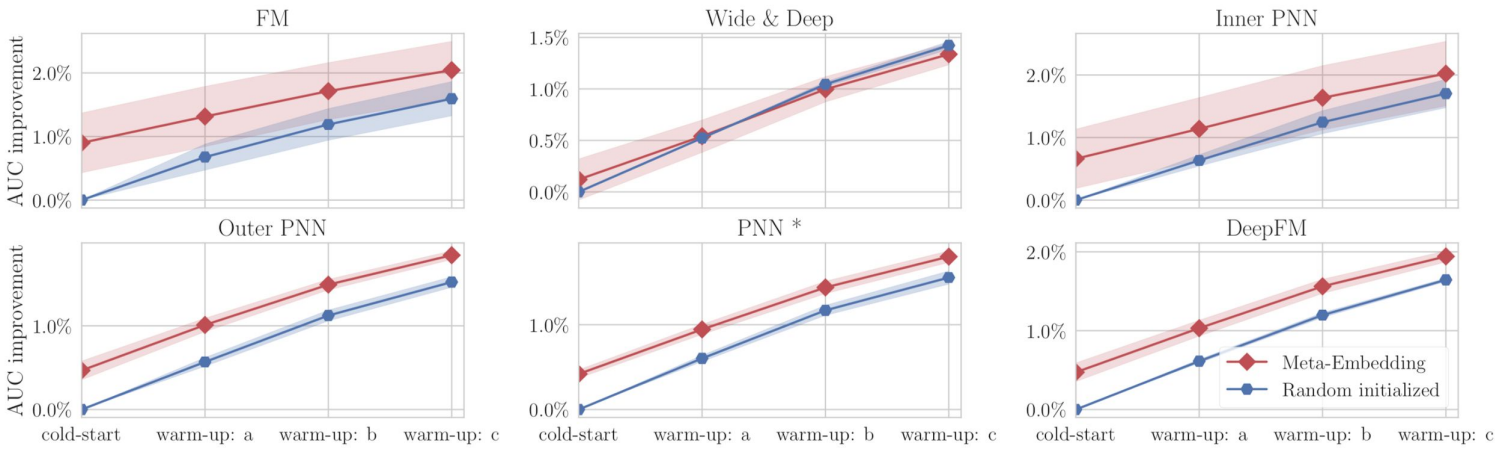
Results: Significantly speed up cold-start phase



The experiment results on small dataset MovieLens.

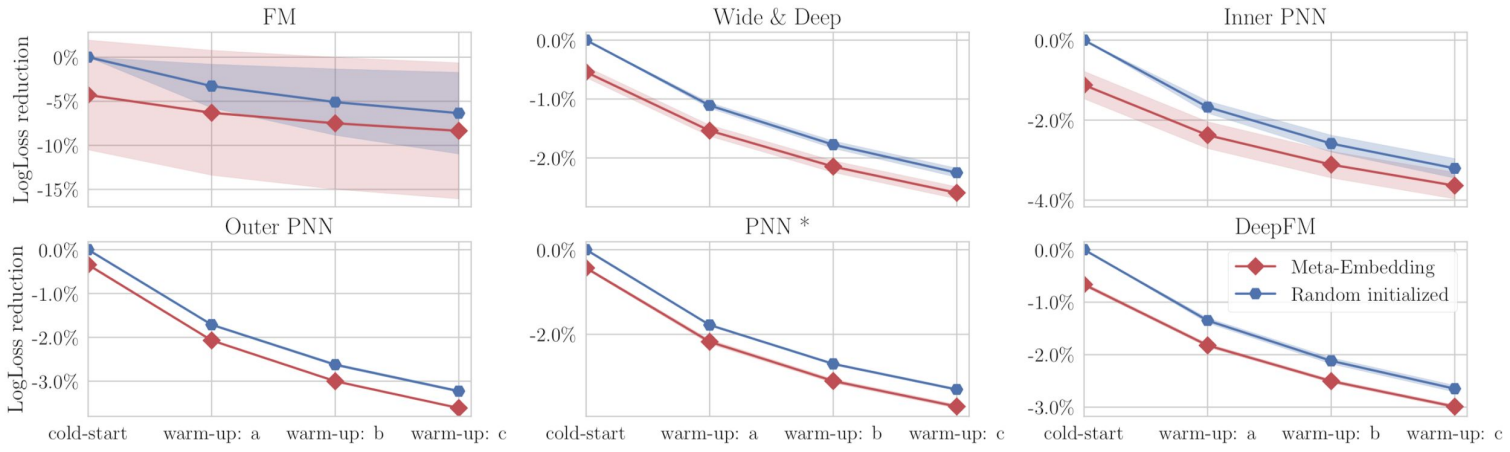
Based on DeepFM, there was an improvement of about 15% against our baseline.

Results: Significantly speed up cold-start phase



The results on the Tencent Social Ads competition 2018 dataset for conversion rate prediction

Results: Significantly speed up cold-start phase



The results on KDD cup 2012 CTR prediction dataset for search ads

On all the tested datasets and base models, Meta-Embedding significantly improves the performance in both the cold-start and the warm-up phase.

Warm Up Cold-Start Advertisements

Improving CTR predictions via Learning to Learn ID embeddings

Feiyang Pan[1], Shuokai Li[1], Xiang Ao[1], Pingzhong Tang[2], Qing He[1]

[1] Institute of Computing Technology, CAS
[2] Tsinghua University

Paper & Code

Thank you!

Q & A