User Behavior Retrieval for Click-Through Rate Prediction

Jiarui Qin, Weinan Zhang, Xin Wu, Jiarui Jin, Yuchen Fang, Yong Yu

July, 2020

Apex Data & Knowledge Management Lab

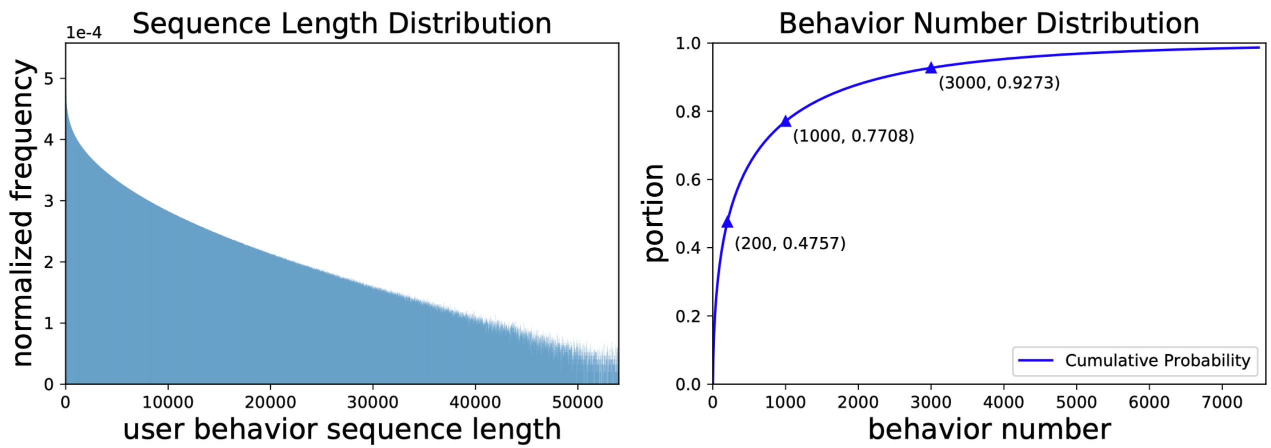
Shanghai Jiao Tong University

Table of Contents

Table of Contents

Background & Motivation

- Users accumulate a large number of behavioral records on the online shopping platforms



- Typical Industrial Solution
  - truncate the sequences and just feed recent behaviors to the prediction model
  - 50 or 100 recent behaviors

K. Ren*, J. Qin*, Y. Fang*, et al. Lifelong Sequential Modeling with Personalized Memorization for User Response Prediction. In SIGIR '19, 2019.

APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

Background & Motivation

- Limitations of current sequential CTR framework
  - Uses the most **recent** N behaviors, if N is large
    - Heavy burden on system overhead (latency + storage), DIEN etc
    - Longer sequences have a lot of noise
    - Each prediction pair uses exactly the same recent N behaviors, the pattern may be not embeded in the recent N behaviors

- Instead of designing more complex model, we turn to the data perspective
  - For each prediction, retrieve the most **useful** N behaviors (N is not large) from all the user's log
  - Use search engine technique

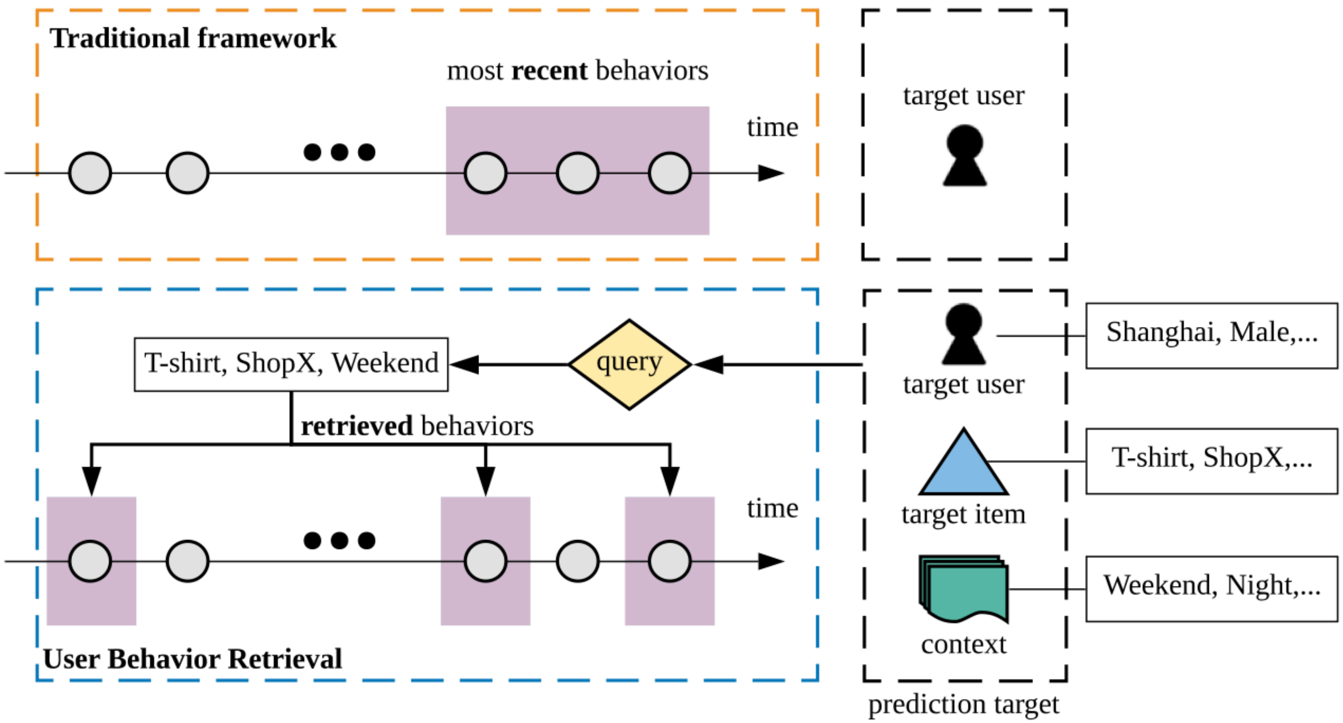
G. Zhou et al. Deep Interest Evolution Network for Click-Through Rate Prediction . In AAAI '19, 2019.

Table of Contents

- Background & Motivation
- Our Methodology
  - Overall Framework
  - User Behavior Retrieval Module
  - Prediction Module
- Training Algorithm
- Experimental Results

Overall Framework

- Differences from original framework:



APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

Overall Framework

- Framework:
  - User Behavior Retrieval Module
  - Prediction Module

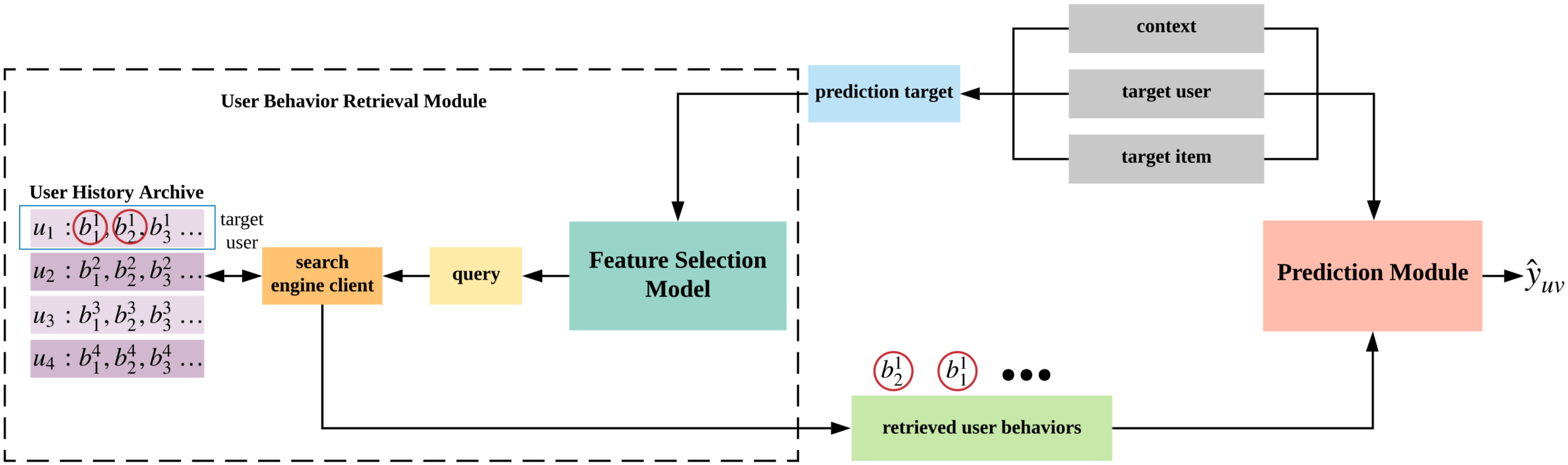


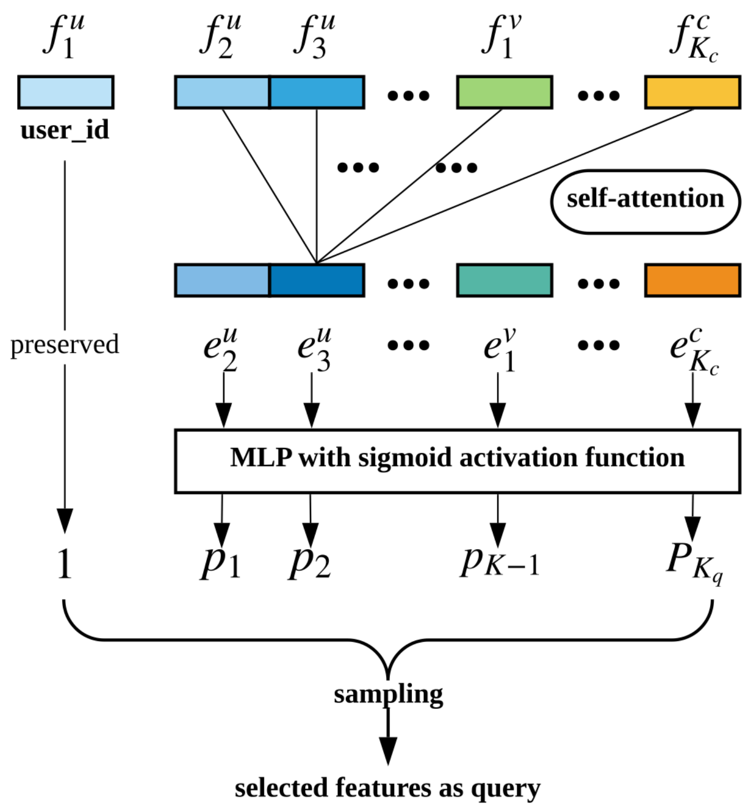
Table of Contents

- Background & Motivation
- Our Methodology
  - Overall Framework
  - User Behavior Retrieval Module
  - Prediction Module
- Training Algorithm
- Experimental Results

APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

User Behavior Retrieval Module

- User behavior retrieval module:
  - Feature selection model using self-attention:



APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

User Behavior Retrieval Module

- User behavior retrieval module:
  - Feature selection model using self-attention:

$$Q = K = V = \begin{pmatrix} f_1 \\ \vdots \\ f_{K_q} \end{pmatrix}$$

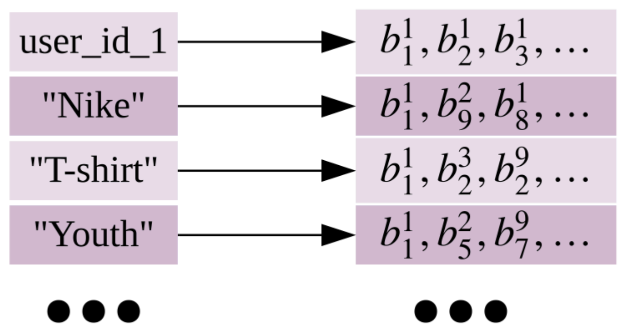
$$\mathrm{SA}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V$$

$$E = \mathrm{Multihead}(Q, K, V) = \mathrm{Concat}(head_1, \ldots, head_h)W^O,$$

$$P = \begin{pmatrix} p_1 \\ \vdots \\ p_{K_q} \end{pmatrix} = \sigma(\hat{E}).$$

User Behavior Retrieval Module

- User behavior retrieval module:
  - Behavior searching:
    - Behavior storage: feature based inverted index



    - The query is formulated as

$$f_1^u \ AND \ (f_1 \ OR \ f_2 \ OR ... OR \ f_n),$$

User Behavior Retrieval for CTR Prediction

- User behavior retrieval module:
  - Behavior searching:
    - Use BM25 to score every behavior document

$$s = \sum_{i=1}^{n} \mathrm{IDF}(f_i) \cdot \frac{tf(f_i, D) \cdot (k_1 + 1)}{tf(f_i, D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{\mathrm{avgdl}}\right)},$$

$$\mathrm{IDF}(f_i) = \log \frac{\mathcal{N} - \mathcal{N}(f_i) + 0.5}{\mathcal{N}(f_i) + 0.5},$$

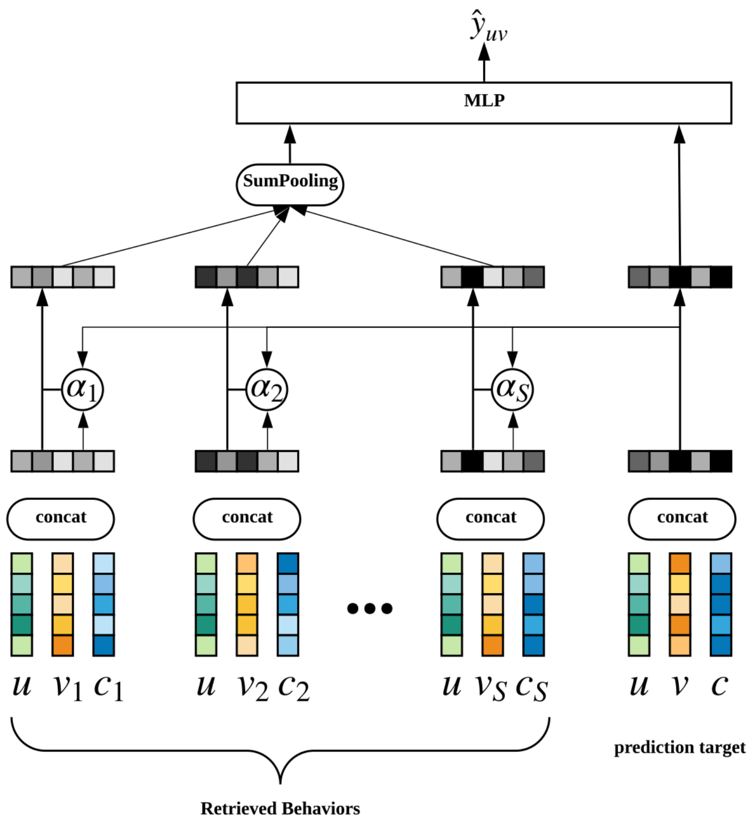
Table of Contents

- Background & Motivation
- Our Methodology
  - Overall Framework
  - User Behavior Retrieval Module
  - Prediction Module
- Training Algorithm
- Experimental Results

APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

Prediction Module

- Prediction module:
  - Attention-based Prediction Network



数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

Table of Contents

Training Algorithm

- Model training:
  - Objective function:

$$J^{\pi_\theta, f_\phi} = \max_\theta \max_\phi \sum_u \sum_v E_{B^u \sim \pi_\theta(B^u|q)}[LL(y_{uv}, f_\phi(B^u, u, v, c))],$$

$$LL(y_{uv}, f_\phi(B^u, u, v, c)) = y_{uv} \cdot \log(f_\phi(B^u, u, v, c)) + (1 - y_{uv}) \cdot \log(1 - f_\phi(B^u, u, v, c)).$$

  - Optimize the prediction model:

$$\begin{aligned}\phi^* &= \arg\min_\phi (L_{ce} + \lambda L_r) \\ &= \arg\min_\phi \sum_u \sum_v E_{B^u \sim \pi_\theta(B^u|q)}[-LL(y_{uv}, f_\phi(B^u, u, v, c))] \\ &+ \frac{1}{2}\lambda \left(\|\Phi\|_2^2 \right),\end{aligned}$$

APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

Training Algorithm

- Model training:
  - As the retrieval operation is discrete so we use policy gradient to estimate the gradient of that operation
  - Optimize the retrieval module:

$$\theta^* = \arg\max_{\theta} \sum_{u} \sum_{v} E_{B^u \sim \pi_\theta(B^u|q)}[LL(y_{uv}, f_\phi(B^u, u, v, c))].$$

$$J^q = E_{B^u \sim \pi_\theta(B^u|q)}[LL(\cdot)]$$

$$\begin{aligned}
\nabla_\theta(J^q) &= \nabla_\theta E_{B^u \sim \pi_\theta(B^u|q)}[LL(\cdot)] \\
&= \sum_{B_i^u \in \mathcal{B}} \nabla_\theta \pi_\theta(B_i^u|q)[LL(\cdot)] \\
&= \sum_{B_i^u \in \mathcal{B}} \pi_\theta(B_i^u|q) \nabla_\theta \log(\pi_\theta(B_i^u|q))[LL(\cdot)] \\
&= E_{B^u \sim \pi_\theta(B^u|q)} \nabla_\theta \log(\pi_\theta(B^u|q))[LL(\cdot)] \\
&\simeq \frac{1}{L} \sum_{l=1}^{L} \nabla_\theta \log(\pi_\theta(B_l^u|q))[LL(\cdot)],
\end{aligned}$$

Training Algorithm

- Model training:

Algorithm 1 Training the UBR4CTR framework

Require: Dataset $\mathcal{D} = (\mathcal{U}_{target}, \mathcal{V}_{target}, \mathcal{C}_{target})$ containing all the target user-item-context triples; User history archive H_u.

Ensure: final CTR prediction $\hat{Y}$ between all the target user u and target item v.

```
Initialize all parameters.
Select the features and form the queries Q = {q, ...} for each prediction target [u, v, c] ∈ D using the initialized feature selection model.
Obtain the retrieved behaviors B = {B^u, ...} of the queries Q using the search engine as described in Section 3.2.2.
Train the attention-based prediction network using Eq. 17 for one epoch.
repeat
    Train retrieval model using Eq. 18 for one epoch.
    Select the features and form the queries Q = {q, ...} for each prediction target [u, v, c] ∈ D using the feature selection model.
    Obtain the retrieved behaviors B = {B^u, ...} of the queries Q using the search engine as described in Section 3.2.2.
    Train attention-based prediction network using Eq. 17 for one epoch.
until convergence
```

Table of Contents

- Background & Motivation
- Our Methodology
  - Overall Framework
  - User Behavior Retrieval Module
  - Prediction Module
- Training Algorithm
- Experimental Results

Experimental Results

- Experimental results:
  - Group1: Same length of user sequences

Model	Tmall		Taobao		Alipay	
	AUC	LL	AUC	LL	AUC	LL
GRU4Rec	0.762	0.585	0.677	0.661	0.6131	0.699
Caser	0.762	0.579	0.673	0.657	0.655	0.676
SASRec	0.755	0.586	0.670	0.658	0.648	0.679
HPMN	0.763	0.579	0.668	0.660	0.615	0.703
MIMN	0.753	0.591	0.662	0.686	0.664	0.675
DIN	0.766	0.576	<u>0.678</u>	<u>0.649</u>	<u>0.732</u>	<u>0.616</u>
DIEN	<u>0.775</u>	<u>0.567</u>	0.677	0.659	0.730	<u>0.616</u>
UBR4CTR	**0.807**	**0.516**	**0.752**	**0.571**	**0.895**	**0.417**
Imprv.	4.1%	9.0%	10.9%	12.0%	22.3%	32.3%

APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

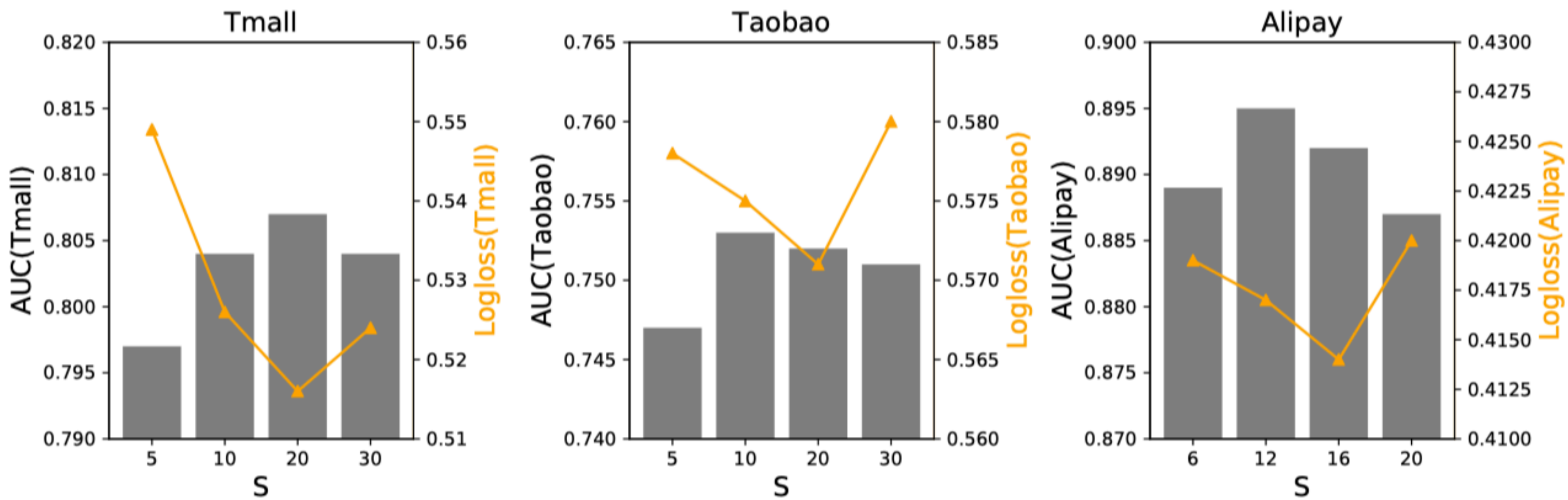
Experimental Results

- Experimental results:
  - Group2: Longer sequences for baselines

Model	Tmall		Taobao		Alipay	
	AUC	LL	AUC	LL	AUC	LL
GRU4Rec	0.781	0.560	0.677	0.660	0.639	0.684
Caser	0.774	0.566	0.645	0.659	0.705	0.631
SASRec	0.769	0.578	0.669	<u>0.654</u>	0.711	0.637
HPMN	0.767	0.579	0.655	0.664	0.703	0.643
MIMN	0.759	0.590	0.659	0.659	0.719	0.634
DIN	0.791	0.546	0.605	0.679	<u>0.856</u>	0.506
DIEN	<u>0.805</u>	<u>0.538</u>	<u>0.704</u>	0.656	0.843	<u>0.491</u>
UBR4CTR	**0.807**	**0.516**	**0.752**	**0.571**	**0.895**	**0.417**
Imprv.	0.2%	4.1%	6.8%	12.7%	4.6%	15.1%

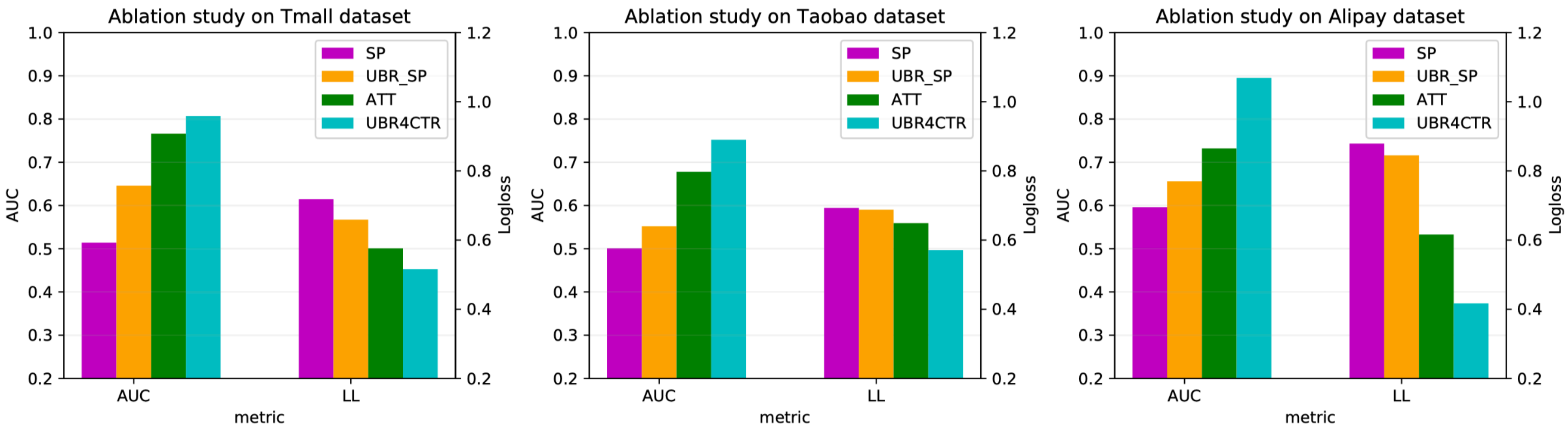
Experimental Results

- Experimental results:
  - Robustness study:



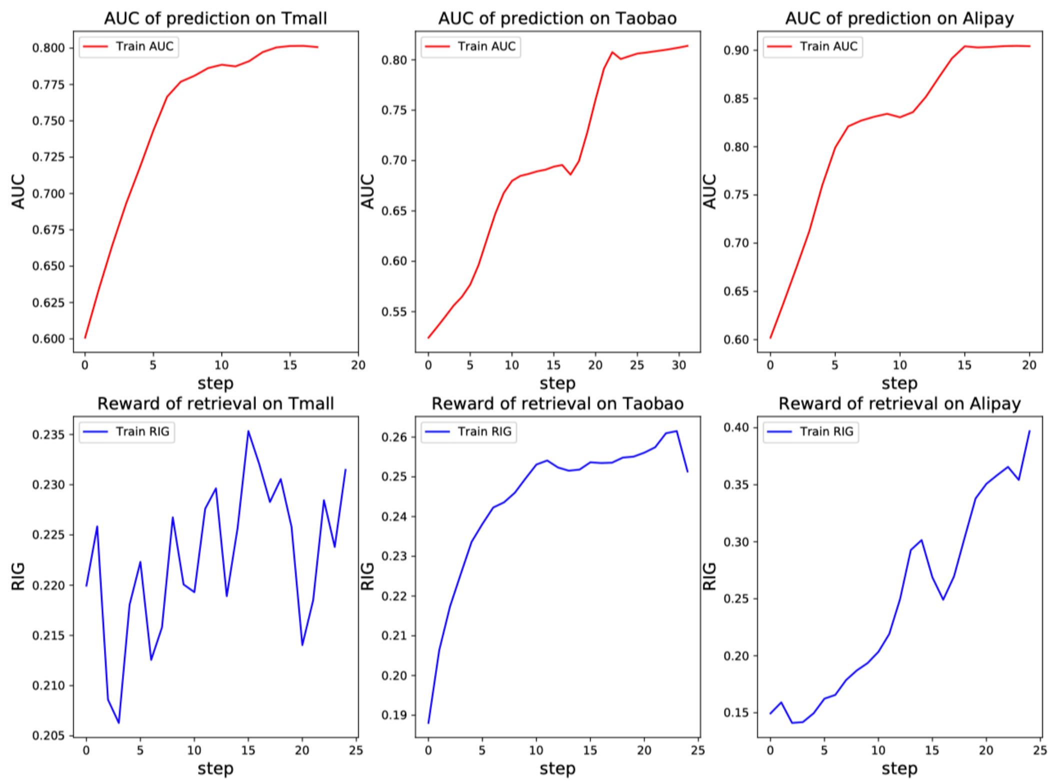
Experimental Results

- Experimental results:
  - Ablation study:



Experimental Results

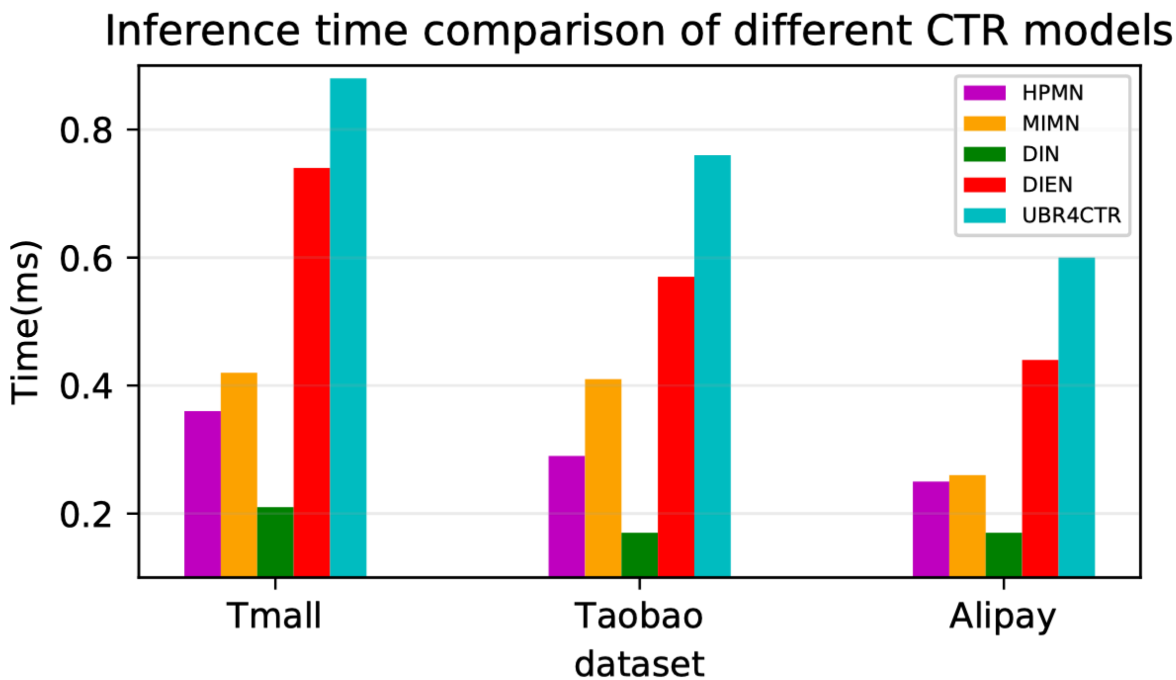
- Experimental results:
  - Learning process:



APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

Experimental Results

- Experimental results:
  - Deployment feasibility:



Paper & Code

- Paper:
  - URL: https://arxiv.org/pdf/2005.14171.pdf
  - QR code:
- Code Repository:
  - URL: https://github.com/qinjr/UBR4CTR
  - QR code: