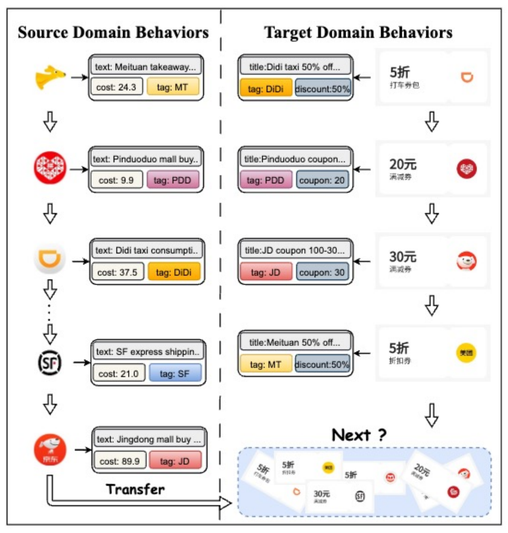
Retrieval Augmented Cross-Domain LifeLong Behavior Modeling for Enhancing Click-through Rate Prediction

Xing Tang, Chaohua Yang, Yuwen Fu, Dongyang Ao, Shiwei Li, Fuyuan Lyu, Dugang Liu and Xiuqiang He

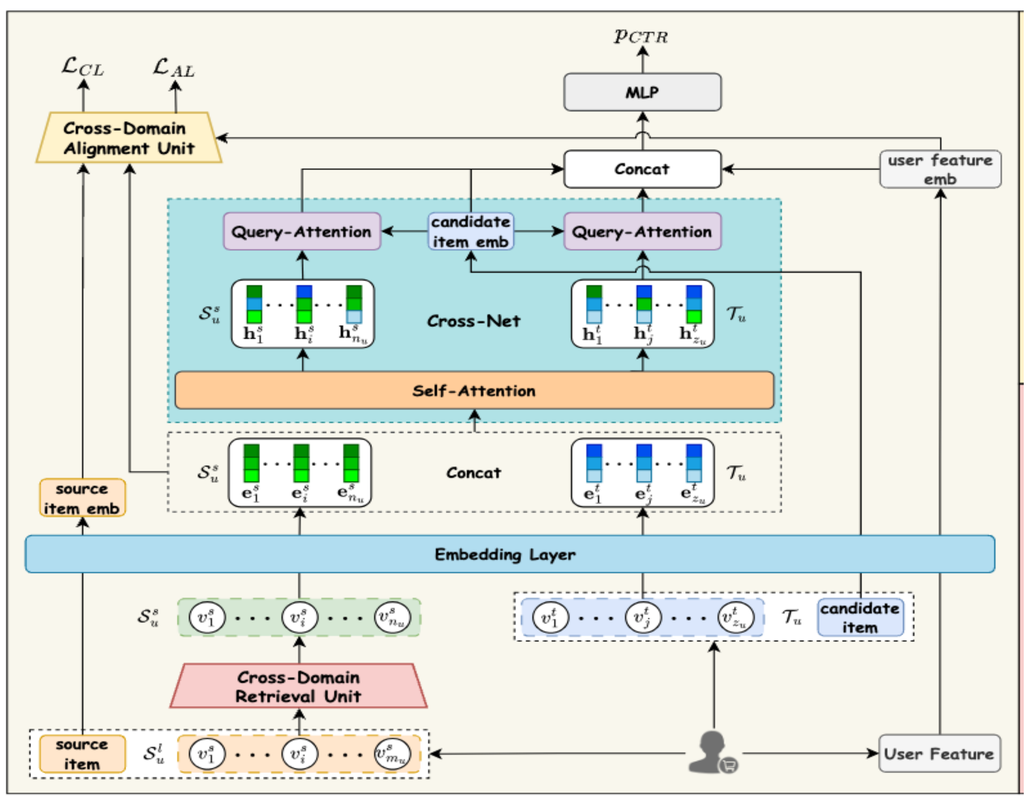
Motivation



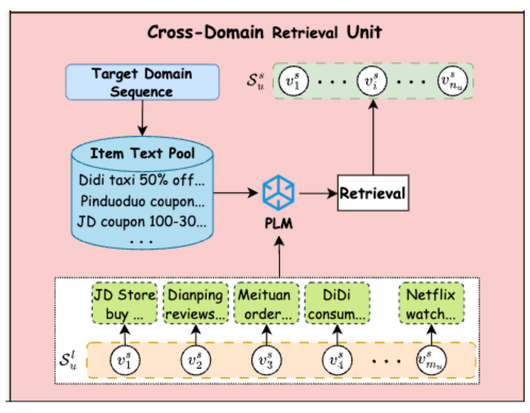
- Lifelong behavior modeling in single domain is an effective way to improve the performance.
- Lifelong behaviors from another domain is helpful.
- Cross-domain lifelong behavior modeling faces inconsistency, sparse and alignment issue.

The RAL-CDNet

- The main framework:



- Cross-domain Retrieval Unit

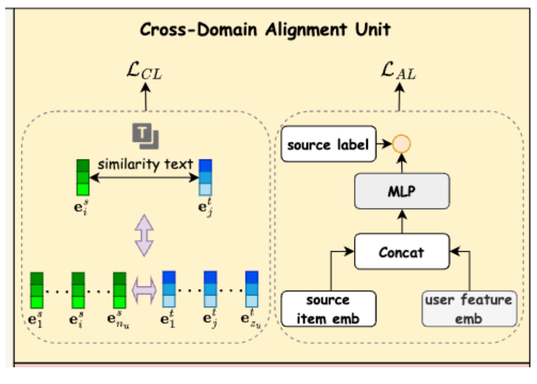


$$m_i = 1 \iff \exists j \in \{1, \cdots, z_u\}, \cos(\text{rep}_i^s, \text{rep}_j^t) > \theta,$$

$$\text{where} \quad \cos(\text{rep}_i^s, \text{rep}^t) = \frac{\text{rep}_i^s \cdot \text{rep}_j^t}{\|\text{rep}_i^s\| \|\text{rep}_j^t\|}.$$

Use PLM to calculate the semantic representation of item

- Cross-domain Alignment Unit



✓ *Contrastive Loss is used to align the embedding across the domains.*

✓ *Auxiliary Loss is used to preserve the source information.*

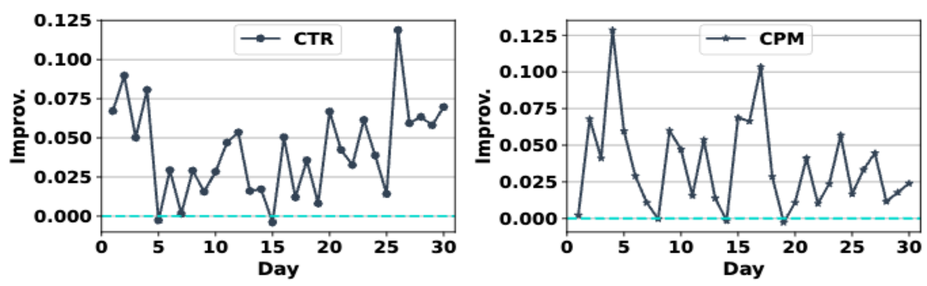
Offline Experiment

Dataset	Amazon				Industrial	
Source → Target	Book → Movie		Book → CD		Payment → Ad	
	AUC↑	Logloss↓	AUC↑	Logloss↓	AUC↑	Logloss↓
DNN	0.7584	0.4413	0.7153	0.3255	0.8361	0.2100
DeepFM	0.7574	0.4421	0.7154	0.3306	0.8371	0.2096
DIN	0.7617	0.4672	0.7164	0.3293	0.8381	0.2091
SIM(Hard)	0.7612	0.5200	0.7158	0.3255	0.8380	0.2091
SIM(Soft)	0.7618	0.5128	0.7163	0.3253	0.8382	0.2091
TWIN	0.7624	0.4690	0.7176	0.3290	0.8379	0.2092
CoNet	0.7635	0.4270	0.7212	**0.3228**	0.8378	0.2093
CDAnet	0.7640	0.4267	0.7215	0.3229	0.8381	0.2094
MiNet	0.7662	0.4311	0.7269	0.3449	0.8384	0.2092
LCN	0.7643	0.4318	0.7221	0.3323	0.8387	0.2089
RAL-CDNet(Hard)	0.7708	0.4230	0.7301	0.3309	0.8405	0.2080
RAL-CDNet	**0.7717***	**0.4210***	**0.7310***	0.3291	**0.8407***	**0.2079**

Observations:

- Lifelong behaviors are useful. *Sequential > DNN/DeepFM*
- Cross-domain methods enhance the performance. *Cross-domain > Single domain*
- Cross-domain Sequential methods are better. *MiNet&LCN > others*
- Cross-domain lifelong behavior modeling improve the performance *RAL-CDNet > others*

Online Experiment &case study



- *Online tests on WeChat advertising platform for 4 weeks*
  CTR+5·34%, CPM+7·67%
- *A case study on Amazon dataset*

Semantic embedding retrieval can discover a broader range of related items across the domains

Summary

- Cross-domain lifelong behaviors modeling is an effective way to improve the performance.
- RAL-CDNet:
  - CD-RU is responsible to retrieval cross-domain items
  - CD-AU introduces two task to explore the relation.
- Evaluation on both online and offline datasets.