Disentangling Long and Short-Term Interests for Recommendation

Yu Zheng[1*], Chen Gao[1†], Jianxin Chang[2], Yanan Niu[2], Yang Song[2], Depeng Jin[1], Yong Li[1]
[1]Beijing National Research Center for Information Science and Technology,
Department of Electronic Engineering, Tsinghua University
[2]Beijing Kuaishou Technology Co., Ltd.

WWW 2022
Code: https://github.com/tsinghua-fib-lab/CLSR

Reported by liang li

Motivation&Solution

Motivation：

- However, since there is no manually annotated label for user interests, existing approaches always follow the paradigm of entangling these two aspects, which may lead to inferior recommendation accuracy and interpretability.

Solution：

In this paper, we propose to disentangle long and short-term interests for recommendation with a contrastive learning framework, CLSR.

Problem Statement

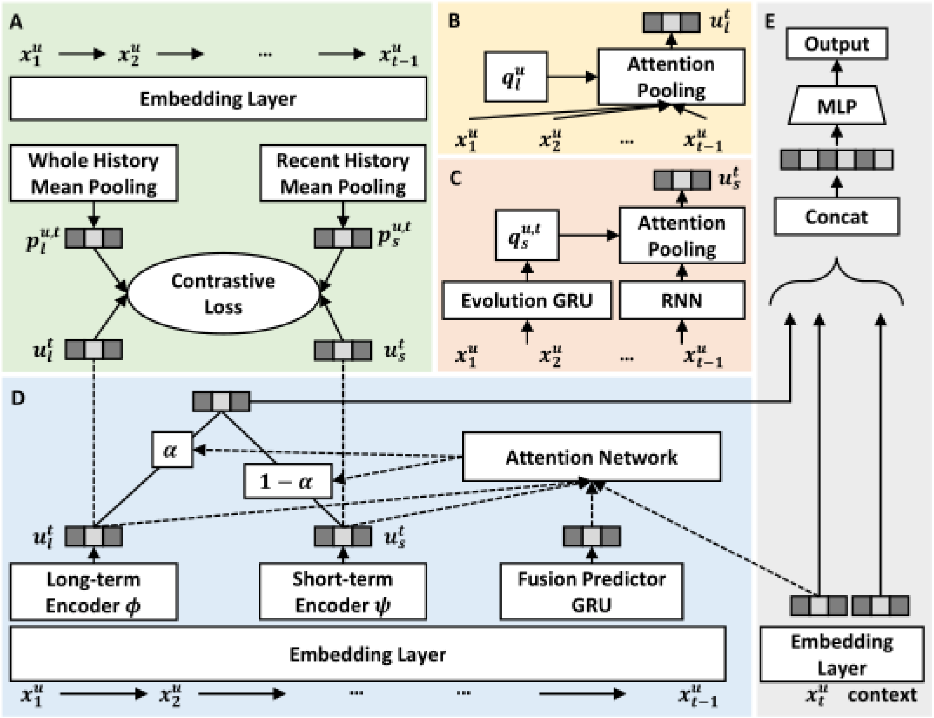

Figure 2: Our proposed CLSR framework based on self-supervised learning. A) contrastive tasks on the similarity between representations and proxies of LS-term interests to enhance disentanglement; B) long-term interests encoder ϕ; C) short-term interests encoder ψ; D) adaptive fusion of LS-term interests with attention on the target item and historical interactions; E) interaction prediction network.

Notations. Let M denote the number of users, and $\{\boldsymbol{x}^u\}_{u=1}^M$ denote the interaction sequences for all users. Each sequence $\boldsymbol{x}^u = [x_1^u, x_2^u, ..., x_{T_u}^u]$ denotes a list of items which are ordered by the corresponding interaction timestamps. Here T_u denotes the length of user u's interaction history, and each item x_t^u is in $[1, N]$, where N denotes the number of items.

Method

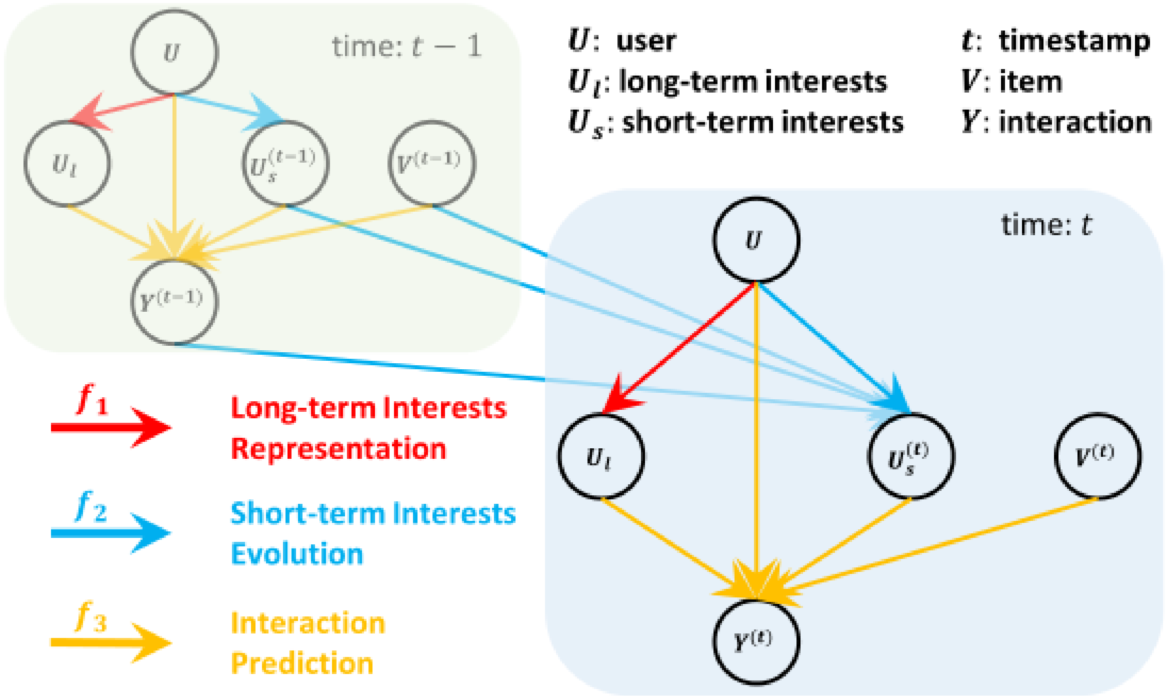

Figure 1: User interests modeling ζ (best viewed in color) which consists of three mechanisms, namely long-term interests representation (red edges), short-term interests evolution (blue edges) and interaction prediction (yellow edges).

$$\zeta = \begin{cases} U_l = f_1(U), & (1) \\ U_s^{(t)} = f_2(U_s^{(t-1)}, V^{(t-1)}, Y^{(t-1)}, U), & (2) \\ Y^{(t)} = f_3(U_l, U_s^{(t)}, V^{(t)}, U), & (3) \end{cases}$$

Method

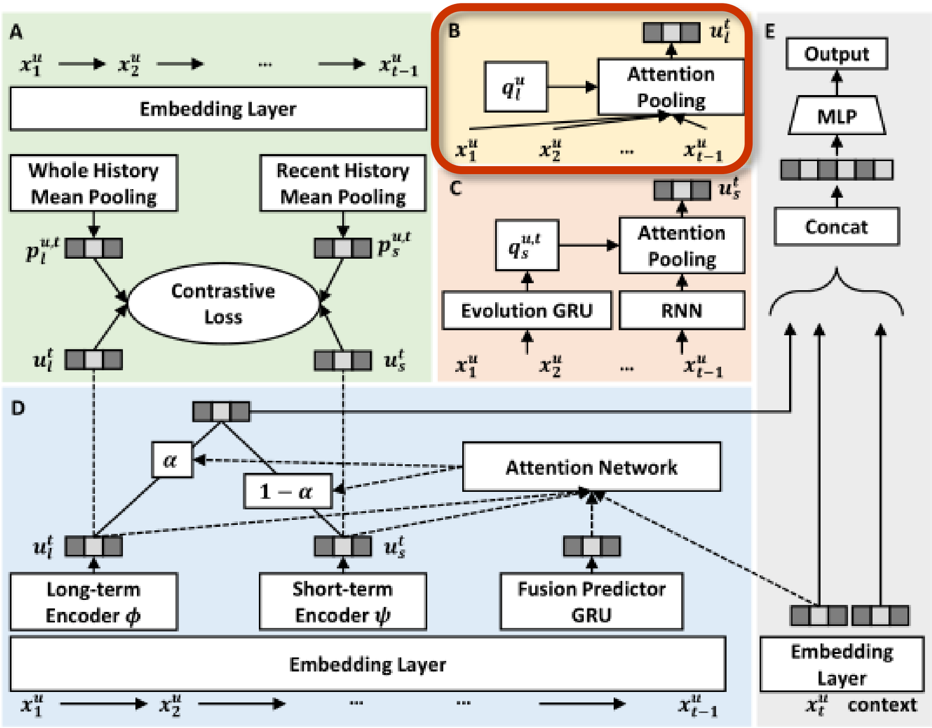

Figure 2: Our proposed CLSR framework based on self-supervised learning. A) contrastive tasks on the similarity between representations and proxies of LS-term interests to enhance disentanglement; B) long-term interests encoder ϕ; C) short-term interests encoder ψ; D) adaptive fusion of LS-term interests with attention on the target item and historical interactions; E) interaction prediction network.

$$q_l^u = \mathrm{Embed}(u), \tag{6}$$

$$u_l^t = \phi(q_l^u, \{x_1^u, \cdots, x_t^u\}), \tag{8}$$

$$v_j = W_l E(x_j^u), \tag{10}$$

$$\alpha_j = \tau_l(v_j \| q_l^u \| (v_j - q_l^u) \| (v_j \cdot q_l^u)), \tag{11}$$

$$a_j = \frac{exp(\alpha_j)}{\sum_{i=1}^{t} exp(\alpha_i)}, \tag{12}$$

$$u_l^t = \sum_{j=1}^{t} a_j \cdot E(x_j^u). \tag{13}$$

Method

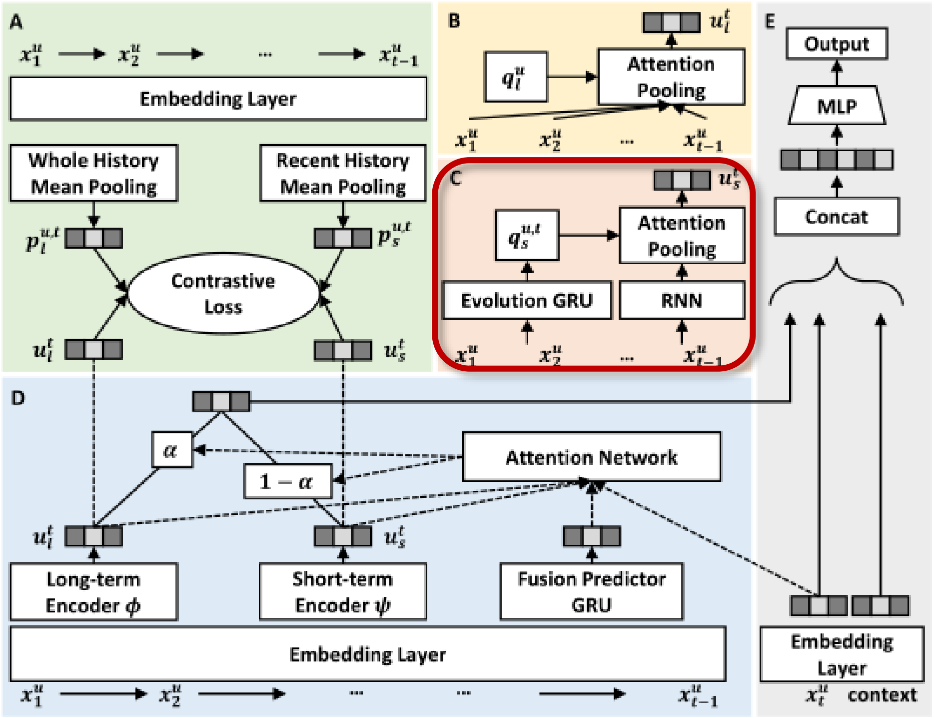

Figure 2: Our proposed CLSR framework based on self-supervised learning. A) contrastive tasks on the similarity between representations and proxies of LS-term interests to enhance disentanglement; B) long-term interests encoder ϕ; C) short-term interests encoder ψ; D) adaptive fusion of LS-term interests with attention on the target item and historical interactions; E) interaction prediction network.

$$q_s^{u,t} = \mathrm{GRU}(\{x_1^u, \cdots, x_t^u\}), \tag{7}$$

$$u_s^t = \psi(q_s^{u,t}, \{x_1^u, \cdots, x_t^u\}), \tag{9}$$

$$\{o_1^u, ..., o_t^u\} = \rho(\{E(x_1^u), ..., E(x_t^u)\}), \tag{14}$$

$$v_j = W_s o_j^u, \tag{15}$$

$$u_s^t = \sum_{j=1}^{t} b_j \cdot o_j^u. \tag{16}$$

Method

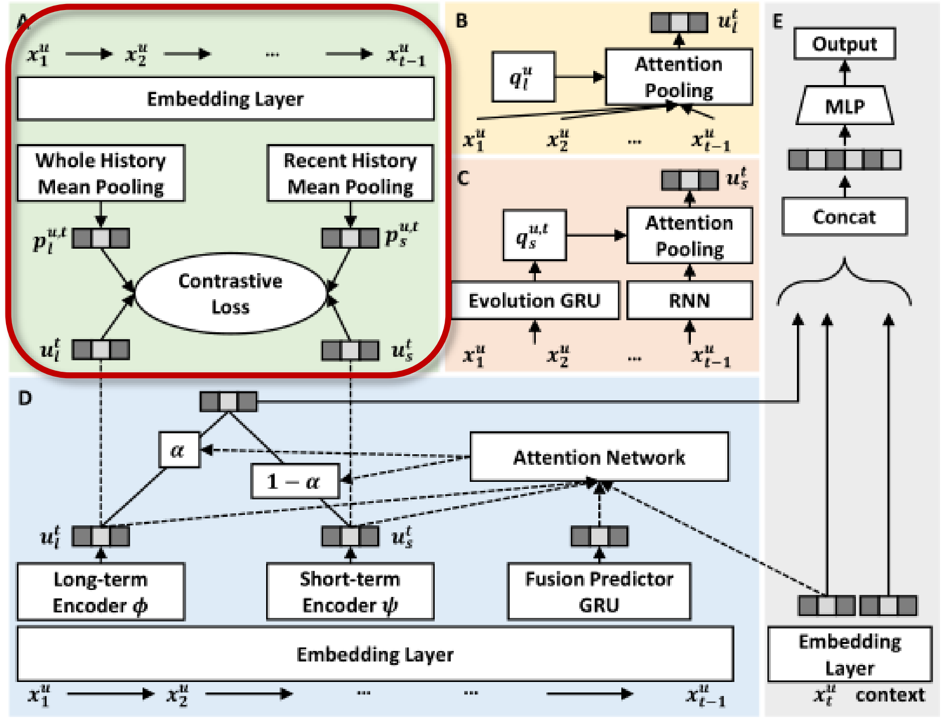

Figure 2: Our proposed CLSR framework based on self-supervised learning. A) contrastive tasks on the similarity between representations and proxies of LS-term interests to enhance disentanglement; B) long-term interests encoder ϕ; C) short-term interests encoder ψ; D) adaptive fusion of LS-term interests with attention on the target item and historical interactions; E) interaction prediction network.

$$p_l^{u,t} = \mathrm{MEAN}(\{x_1^u, \cdots, x_t^u\}) = \frac{1}{t}\sum_{j=1}^{t} E(x_j^u), \tag{17}$$

$$p_s^{u,t} = \mathrm{MEAN}(\{x_{t-k+1}^u, \cdots, x_t^u\}) = \frac{1}{k}\sum_{j=1}^{k} E(x_{t-j+1}^u), \tag{18}$$

$$sim(u_l^t, p_l^{u,t}) > sim(u_l^t, p_s^{u,t}), \tag{19}$$

$$sim(p_l^{u,t}, u_l^t) > sim(p_l^{u,t}, u_s^t), \tag{20}$$

$$sim(u_s^t, p_s^{u,t}) > sim(u_s^t, p_l^{u,t}), \tag{21}$$

$$sim(p_s^{u,t}, u_s^t) > sim(p_s^{u,t}, u_l^t), \tag{22}$$

$$\mathcal{L}_{\mathrm{bpr}}(a, p, q) = \sigma(\langle a, q\rangle - \langle a, p\rangle), \tag{23}$$

$$\mathcal{L}_{\mathrm{tri}}(a, p, q) = \max\{d(a, p) - d(a, q) + m, 0\}, \tag{24}$$

$$\mathcal{L}_{\mathrm{con}}^{u,t} = f(u_l, p_l, p_s) + f(p_l, u_l, u_s) + f(u_s, p_s, p_l) + f(p_s, u_s, u_l) \tag{25}$$

where we omit the superscript of interest representations and proxies, and f can be either $\mathcal{L}_{bpr}$ or $\mathcal{L}_{tri}$.

Method

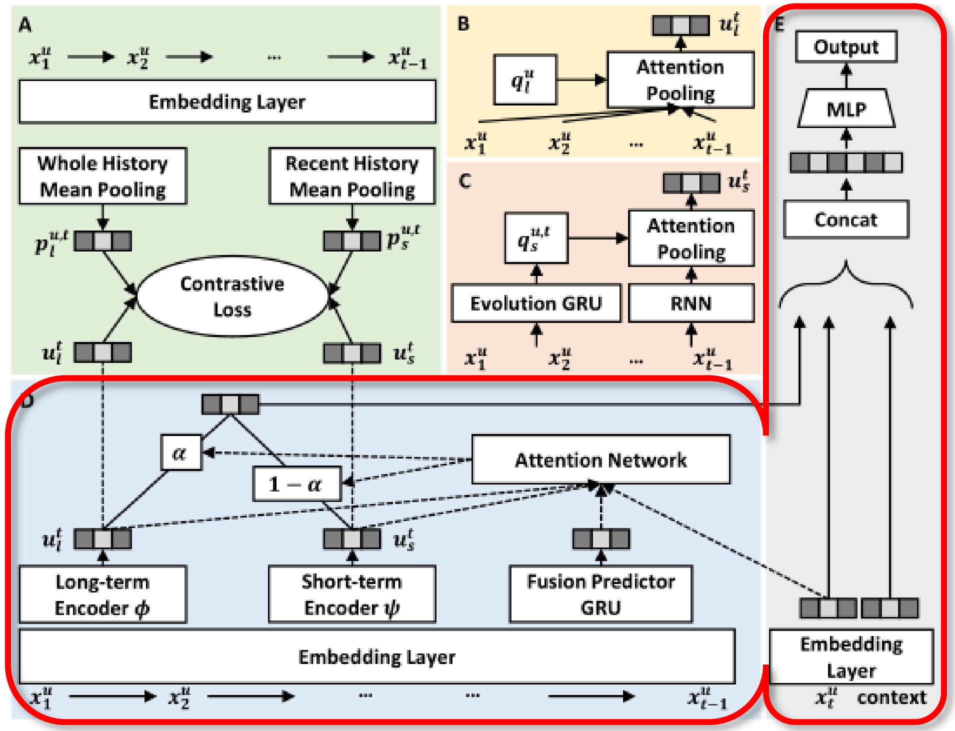

Figure 2: Our proposed CLSR framework based on self-supervised learning. A) contrastive tasks on the similarity between representations and proxies of LS-term interests to enhance disentanglement; B) long-term interests encoder ϕ; C) short-term interests encoder ψ; D) adaptive fusion of LS-term interests with attention on the target item and historical interactions; E) interaction prediction network.

$$\boldsymbol{h}_t^u = \mathrm{GRU}(\{E(x_1^u), ..., E(x_t^u)\}), \tag{26}$$

$$\alpha = \sigma(\tau_f(\boldsymbol{h}_t^u \| E(x_{t+1}^u) \| \boldsymbol{u}_l^t \| \boldsymbol{u}_s^t), \tag{27}$$

$$\boldsymbol{u}^t = \alpha \cdot \boldsymbol{u}_l^t + (1-\alpha) \cdot \boldsymbol{u}_s^t, \tag{28}$$

$$\hat{y}_{u,v}^{t+1} = \mathrm{MLP}(\boldsymbol{u}^t \| E(v)). \tag{29}$$

$$\mathcal{L}_{\mathrm{rec}}^{u,t} = -\frac{1}{N} \sum_{v \in O} y_{u,v}^{t+1} \log(\hat{y}_{u,v}^{t+1}) + (1 - y_{u,v}^{t+1}) \log(1 - \hat{y}_{u,v}^{t+1}), \tag{30}$$

$$\mathcal{L} = \sum_{u=1}^{M} \sum_{t=1}^{T_u} (\mathcal{L}_{\mathrm{rec}}^{u,t} + \beta \mathcal{L}_{\mathrm{con}}^{u,t}) + \lambda \|\Theta\|_2, \tag{31}$$

Experiments

Table 1: Statistics of the two datasets used in experiments.

Dataset	Users	Items	Instances	Average Length
Taobao	36,915	64,138	1,471,155	39.85
Kuaishou	60,813	292,286	14,952,659	245.88

Table 2: Overall performance on Taobao and Kuaishou datasets. Underline means the best two baselines, bold means p-value < 0.05, * means p-value < 0.01, and ** means p-value < 0.001.

Dataset		Taobao				Kuaishou			
Category	Method	AUC	GAUC	MRR	NDCG@2	AUC	GAUC	MRR	NDCG@2
Long-term	NCF	0.7128	0.7221	0.1446	0.0829	0.5559	0.5531	0.7734	0.8327
	DIN	0.7637	0.8524	0.3091	0.2352	0.6160	0.7483	0.8863	0.9160
	LightGCN	0.7483	0.7513	0.1669	0.1012	0.6403	0.6407	0.8175	0.8653
Short-term	Caser	0.8312	0.8499	0.3508	0.2890	0.7795	0.8097	0.9100	0.9336
	GRU4REC	0.8635	0.8680	0.3993	<u>0.3422</u>	0.8156	<u>0.8298</u>	<u>0.9166</u>	<u>0.9384</u>
	DIEN	0.8477	<u>0.8745</u>	<u>0.4011</u>	0.3404	0.7037	0.7800	0.9030	0.9284
	SASRec	0.8598	0.8635	0.3915	0.3340	<u>0.8199</u>	0.8293	0.9161	0.9380
	SURGE	<u>0.8906</u>	<u>0.8888</u>	<u>0.4228</u>	<u>0.3625</u>	<u>0.8525</u>	<u>0.8610</u>	<u>0.9316</u>	<u>0.9495</u>
LS-term	SLi-Rec	<u>0.8664</u>	0.8669	0.3617	0.2971	0.7978	0.8128	0.9075	0.9318
	Ours	**0.8953****	**0.8936****	**0.4372****	**0.3788****	**0.8563**	**0.8718**	**0.9382***	**0.9544***

Experiments

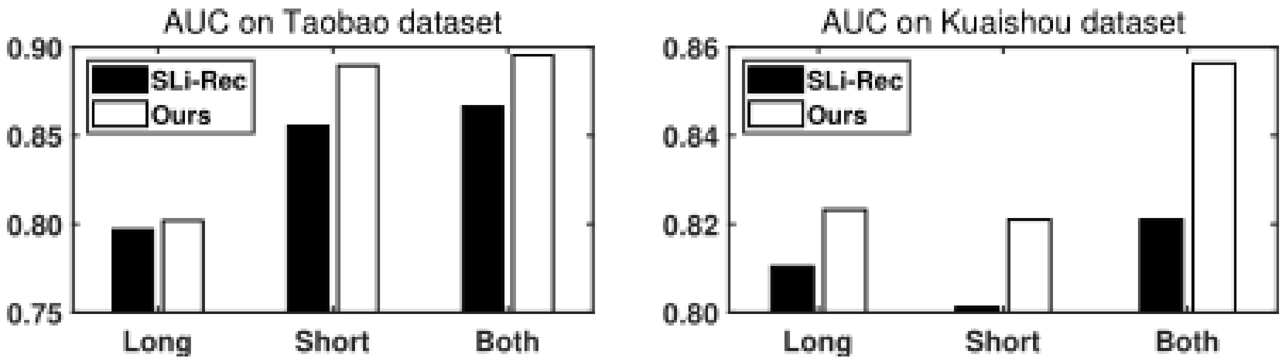

Figure 3: Comparison of using single and both interests between CLSR and Sli-Rec.

Table 3: Comparison between CLSR and SLi-Rec on predicting click and purchase/like.

Dataset	Method	Click		Purchase/Like	
		AUC	AVG(α)	AUC	AVG(α)
Taobao	SLi-Rec	0.8572	0.4651	0.8288	0.4350 (-6.47%)
	CLSR	0.8885	0.3439	0.8616	0.3568 (**+3.75%**)
Kuaishou	SLi-Rec	0.8153	0.7259	0.7924	0.7543 (+3.91%)
	CLSR	0.8618	0.2528	0.7946	0.2757 (**+9.06%**)

On Taobao dataset, α of CLSR for purchase behavior is larger than click by about 4%. However, for SLi-Rec, α for purchase is even less than click by over 6%. On Kuaishou dataset, though α for like is larger than click in both SLi-Rec and CLSR, the relative increment of α for CLSR is over two times larger than SLi-Rec (+9.06% v.s. +3.91%).

Experiments

Table 4: Counterfactual evaluation under shuffle protocol.

Dataset	Method	Click		Purchase/Like	
		AUC	MRR	AUC	MRR
Taobao	SLi-Rec	0.8092	0.2292	0.8480	0.3151
	CLSR	**0.8413**	**0.2744**	**0.8790**	**0.4194**
Kuaishou	SLi-Rec	0.7992	0.9088	0.8165	0.9113
	CLSR	**0.8431**	**0.9380**	**0.8197**	**0.9167**

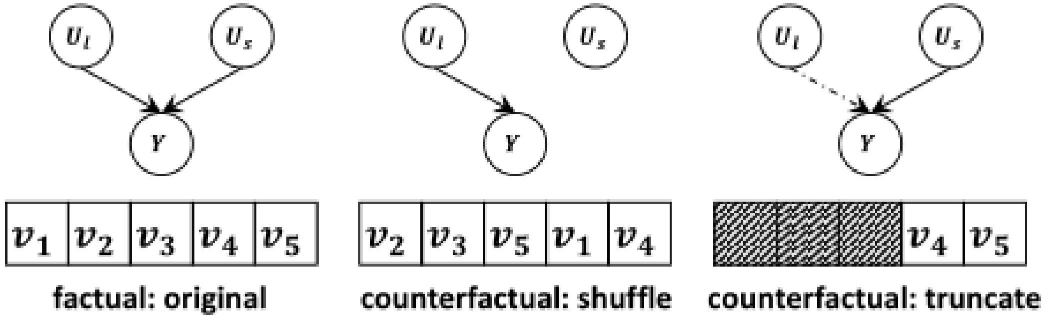

Figure 4: Counterfactual evaluation. Shuffle: short-term interests are removed by shuffling. Truncate: long-term interests are weakened by discarding early history.

Experiments

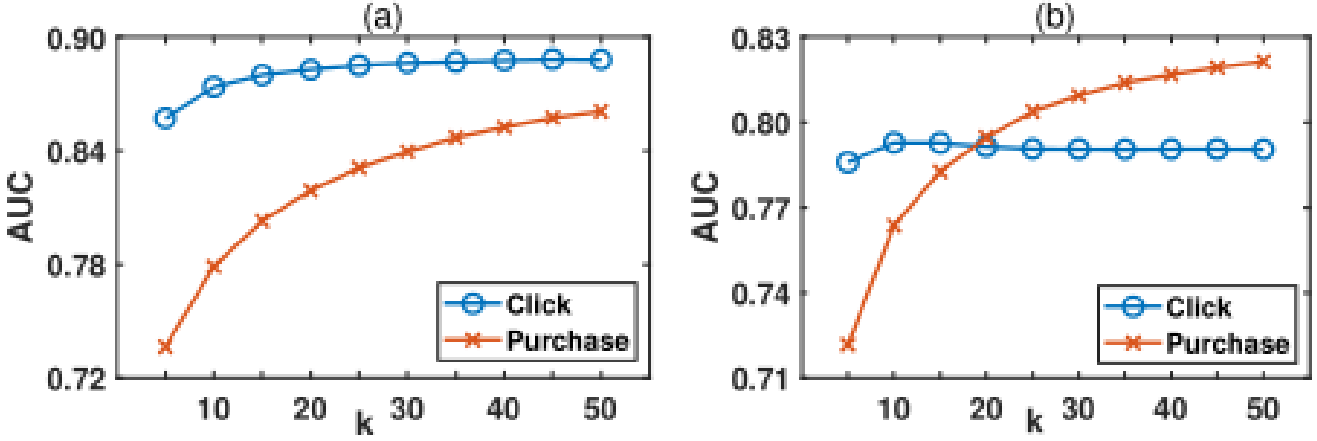

Figure 5: Counterfactual evaluation under truncate protocol. (a) CLSR. (b) CLSR with only long-term interests.

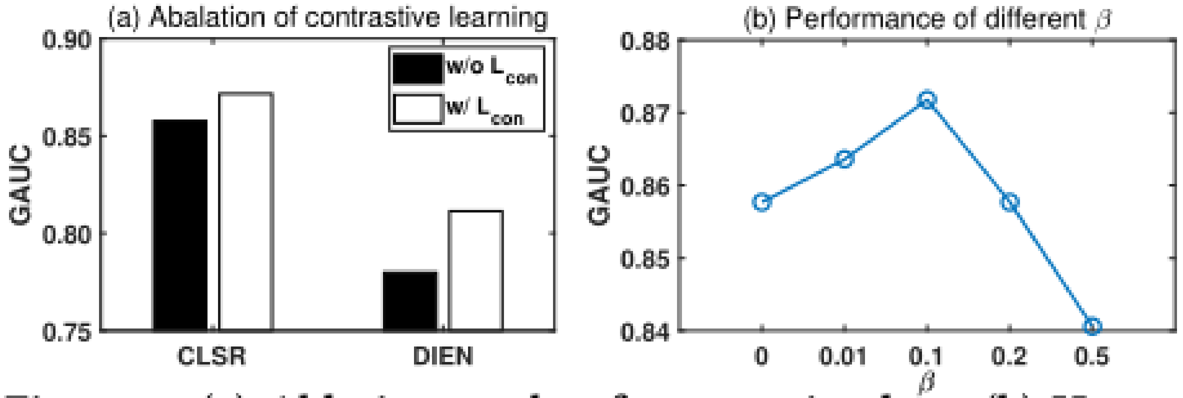

Figure 6: (a) Ablation study of contrastive loss. (b) Hyper-parameter study of β.

Experiments

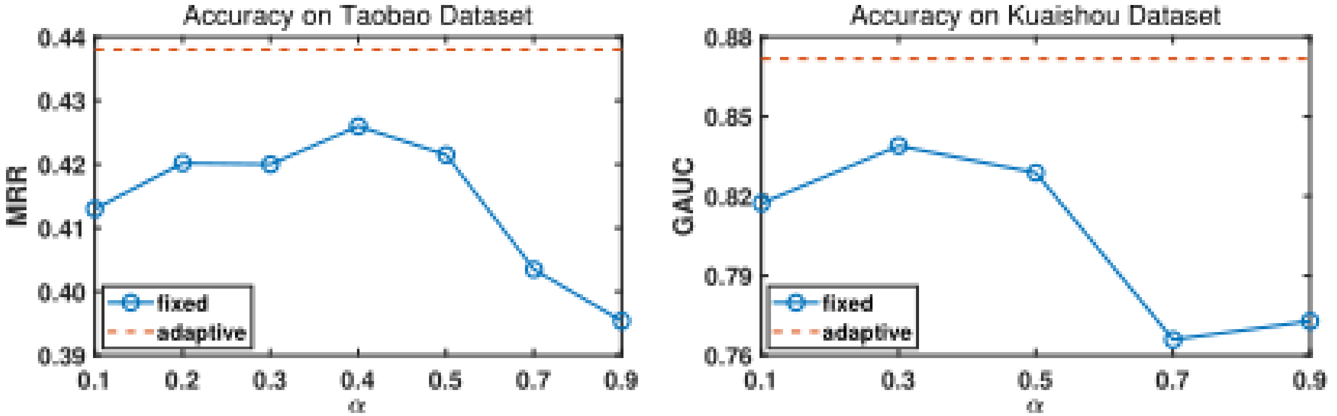

Figure 7: Comparison between adaptive and fixed fusion.

Thanks