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.

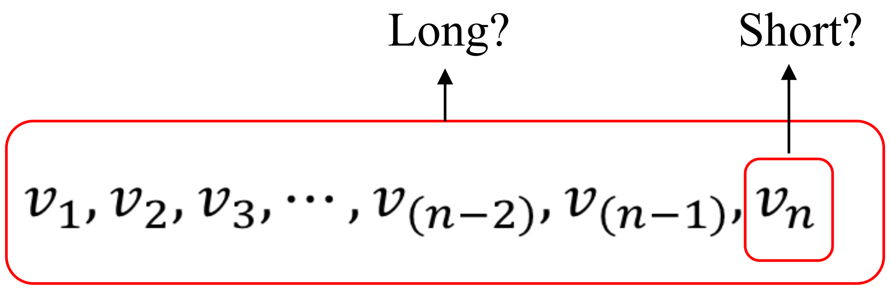
WWW 2022

2022. 3. 22 • ChongQing

Reported by Gu Tang

gesis
Leibniz-Institut
für Sozialwissenschaften

Introduction



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

case as a toy example, suppose a recommendation model entangles LS-term interests as follows,

$$U_l' = 0.6U_l + 0.4U_s, \ U_s' = 0.4U_l + 0.6U_s, \quad (4)$$

where U_l' and U_s' are the learned entangled interests. Given the fusion weights (importance) of LS-term interests as 0.8 and 0.2 respectively, the actual fused interests are computed as follows,

$$U_{fuse}' = 0.8U_l' + 0.2U_s' = 0.56U_l + 0.44U_s, \quad (5)$$

which is quite different from the desired interests.

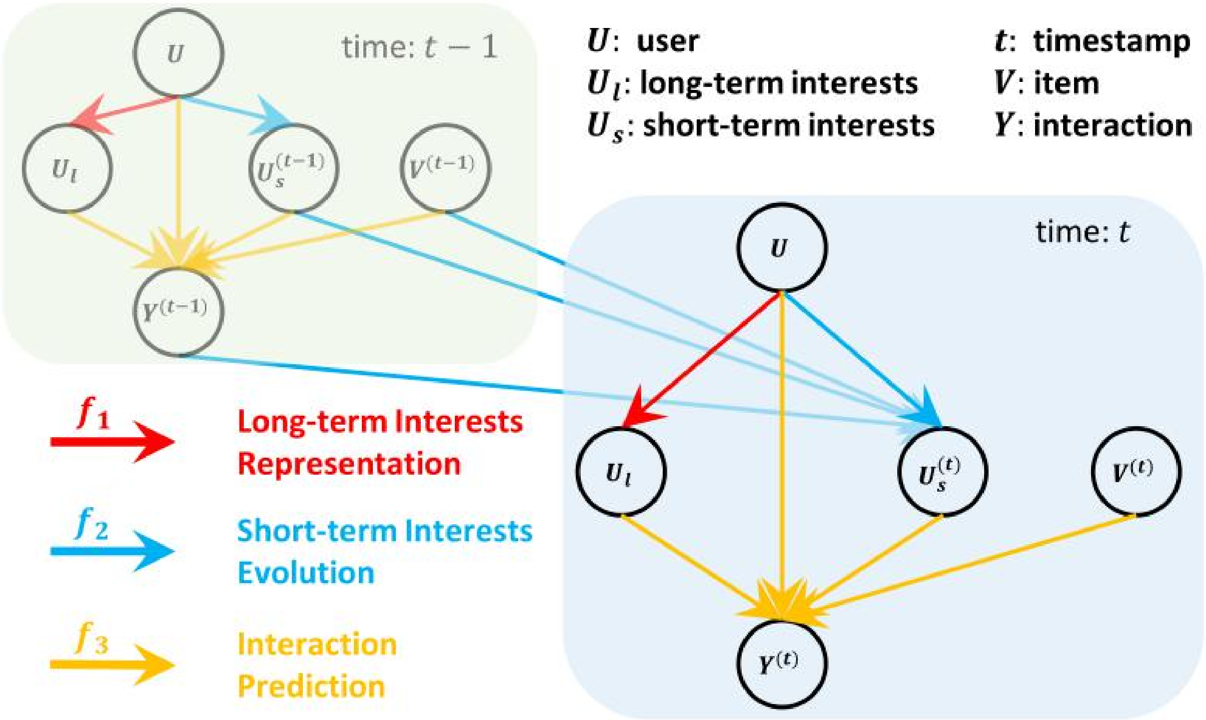


Figure 1: User interests modeling ζ (best viewed in color) which consists of three mechanisms, namely long-term interests representation (red edges), short-term interests evolution (blue edges) and interaction prediction (yellow edges).

Method

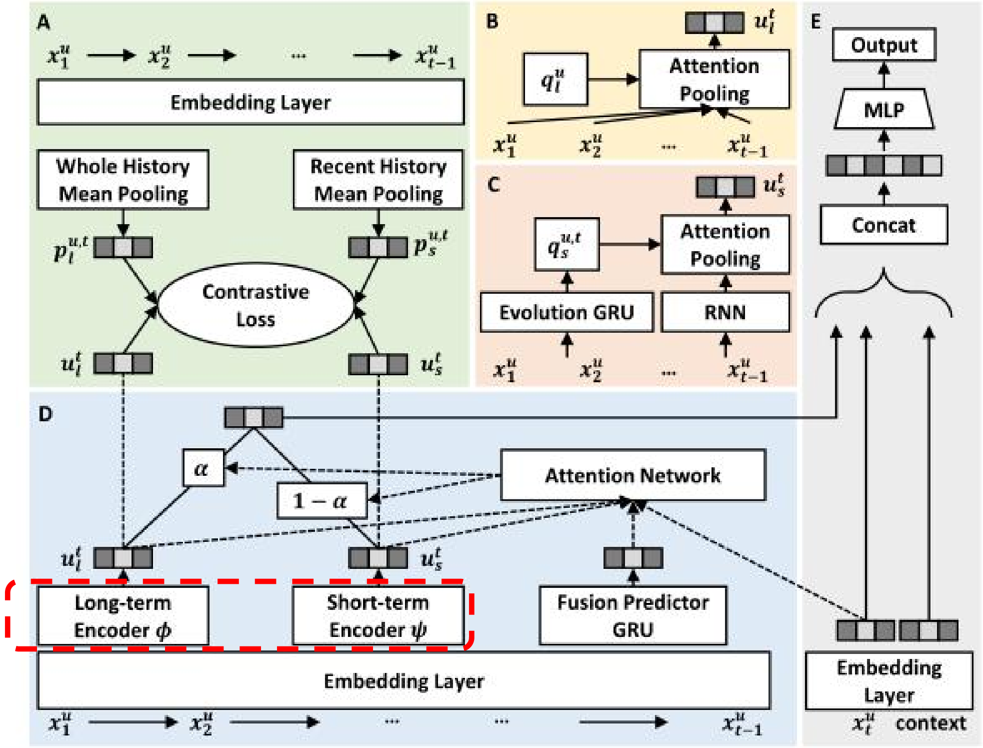


Figure 2: Our proposed CLSR framework based on self-supervised learning. A) contrastive tasks on the similarity between representations and proxies of LS-term interests to enhance disentanglement; B) long-term interests encoder ϕ; C) short-term interests encoder ψ; D) adaptive fusion of LS-term interests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

ρ represents a RNN model.

τ_l is a multi-layer perceptrons　　τ_l 映射到一维

$$\boldsymbol{q}_l^u = \mathrm{Embed}(u), \tag{6}$$

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

$$\boldsymbol{u}_l^t = \phi(\boldsymbol{q}_l^u, \{x_1^u, \cdots, x_t^u\}), \tag{8}$$

$$\boldsymbol{u}_s^t = \psi(\boldsymbol{q}_s^{u,t}, \{x_1^u, \cdots, x_t^u\}), \tag{9}$$

$$v_j = \boldsymbol{W}_l \boldsymbol{E}(x_j^u), \tag{10}$$

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

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

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

$$\{\boldsymbol{o}_1^u, ..., \boldsymbol{o}_t^u\} = \rho(\{\boldsymbol{E}(x_1^u), ..., \boldsymbol{E}(x_t^u)\}), \tag{14}$$

$$v_j = \boldsymbol{W}_s \boldsymbol{o}_j^u, \tag{15}$$

Time4LSTM [47]. Similar as Eqn (10) ~~and~~ to (13), we use $\boldsymbol{q}_s^{u,t}$ as the query vector, and obtain attention scores b_k. Then the learned representation for short-term interests can be computed as follows,

$$\boldsymbol{u}_s^t = \sum_{j=1}^t b_j \cdot \boldsymbol{o}_j^u.$$

Method

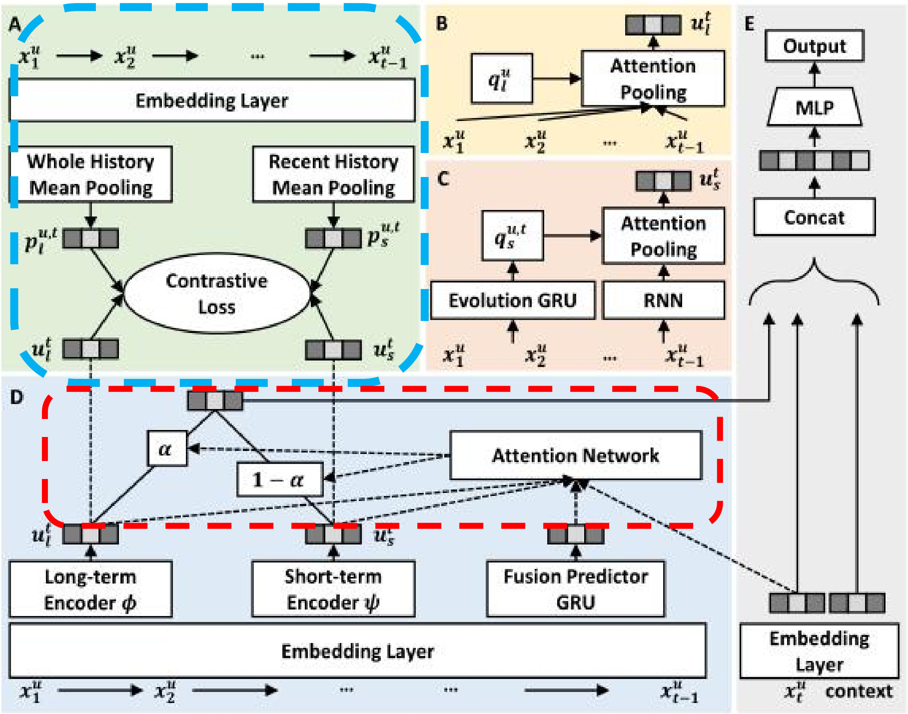


Figure 2: Our proposed CLSR framework based on self-supervised learning. A) contrastive tasks on the similarity between representations and proxies of LS-term interests to enhance disentanglement; B) long-term interests encoder ϕ; C) short-term interests encoder ψ; D) adaptive fusion of LS-term interests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

denotes a positive margin value. Both $\mathcal{L}_{bpr}$ and $\mathcal{L}_{tri}$ are designed for making the anchor a more similar to the positive sample p than the negative sample q. Thus the contrastive loss for self-supervised

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

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

$$sim(\boldsymbol{u}_l^t, \boldsymbol{p}_l^{u,t}) > sim(\boldsymbol{u}_l^t, \boldsymbol{p}_s^{u,t}), \tag{19}$$

$$sim(\boldsymbol{p}_l^{u,t}, \boldsymbol{u}_l^t) > sim(\boldsymbol{p}_l^{u,t}, \boldsymbol{u}_s^t), \tag{20}$$

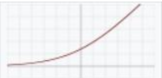
$$sim(\boldsymbol{u}_s^t, \boldsymbol{p}_s^{u,t}) > sim(\boldsymbol{u}_s^t, \boldsymbol{p}_l^{u,t}), \tag{21}$$

$$sim(\boldsymbol{p}_s^{u,t}, \boldsymbol{u}_s^t) > sim(\boldsymbol{p}_s^{u,t}, \boldsymbol{u}_l^t), \tag{22}$$

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

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

$$\mathcal{L}_{\mathrm{con}}^{u,t} = f(\boldsymbol{u}_l, \boldsymbol{p}_l, \boldsymbol{p}_s) + f(\boldsymbol{p}_l, \boldsymbol{u}_l, \boldsymbol{u}_s) + f(\boldsymbol{u}_s, \boldsymbol{p}_s, \boldsymbol{p}_l) + f(\boldsymbol{p}_s, \boldsymbol{u}_s, \boldsymbol{u}_l) \tag{25}$$

SoftPlus[20]  $f(x) = \ln(1 + e^x)$

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

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

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

Method

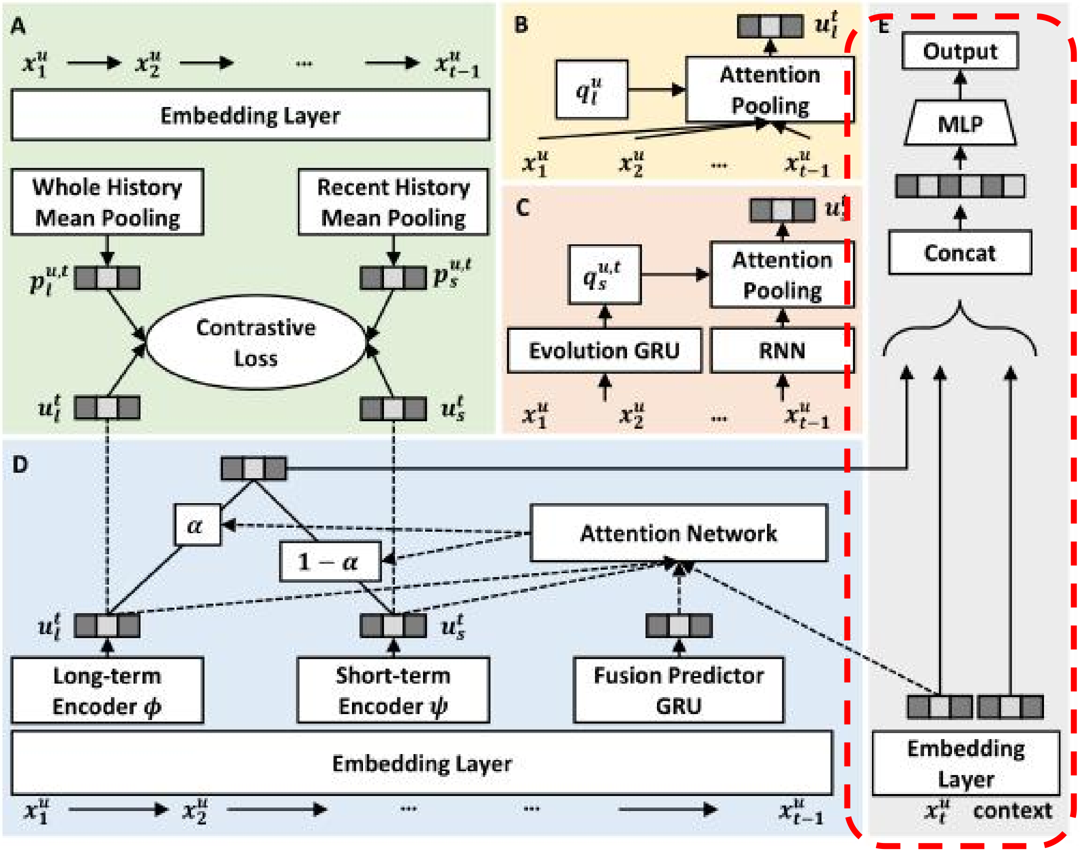


Figure 2: Our proposed CLSR framework based on self-supervised learning. A) contrastive tasks on the similarity between representations and proxies of LS-term interests to enhance disentanglement; B) long-term interests encoder ϕ; C) short-term interests encoder ψ; D) adaptive fusion of LS-term interests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

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

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

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

Method

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***

Experiment

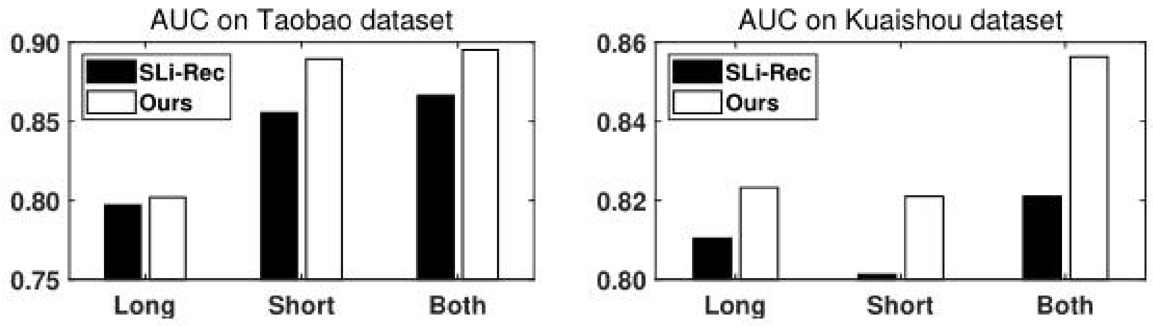


Figure 3: Comparison of using single and both interests between CLSR and Sli-Rec.

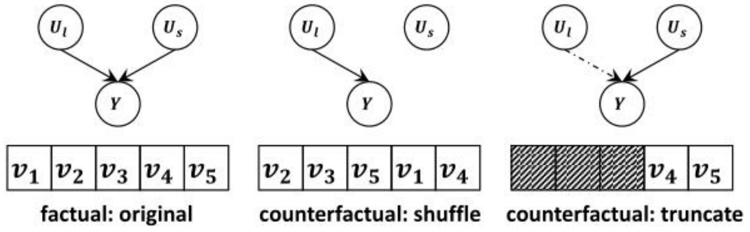


Figure 4: Counterfactual evaluation. Shuffle: short-term interests are removed by shuffling. Truncate: long-term interests are weakened by discarding early history.

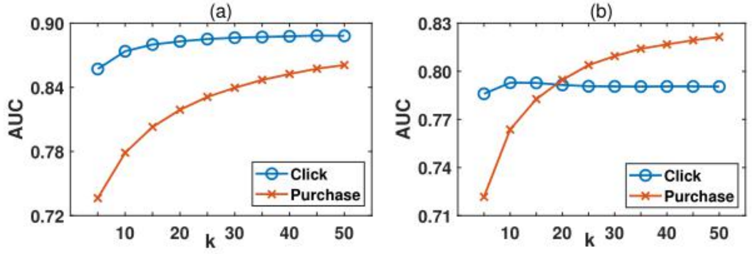


Figure 5: Counterfactual evaluation under truncate protocol. (a) CLSR. (b) CLSR with only long-term interests.

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%)**

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**

Experiments

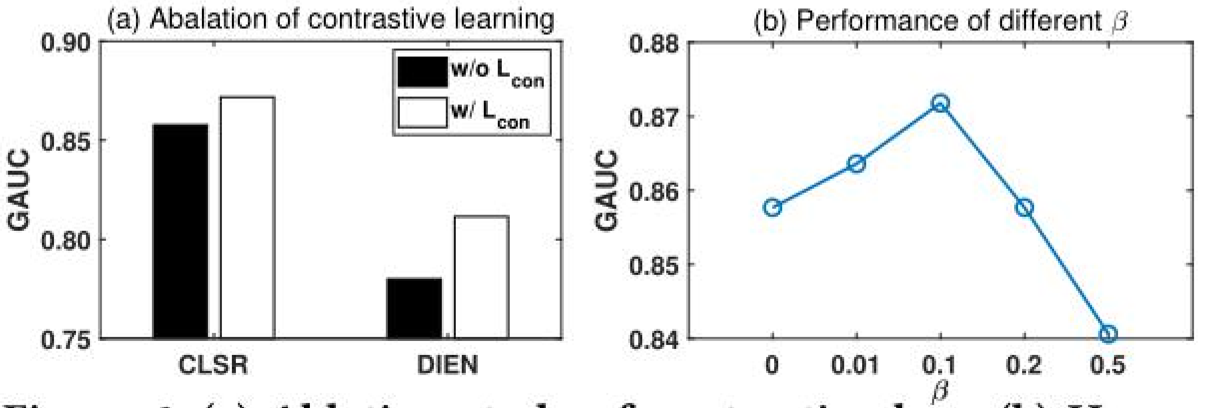


Figure 6: (a) Ablation study of contrastive loss. (b) Hyper-parameter study of β.

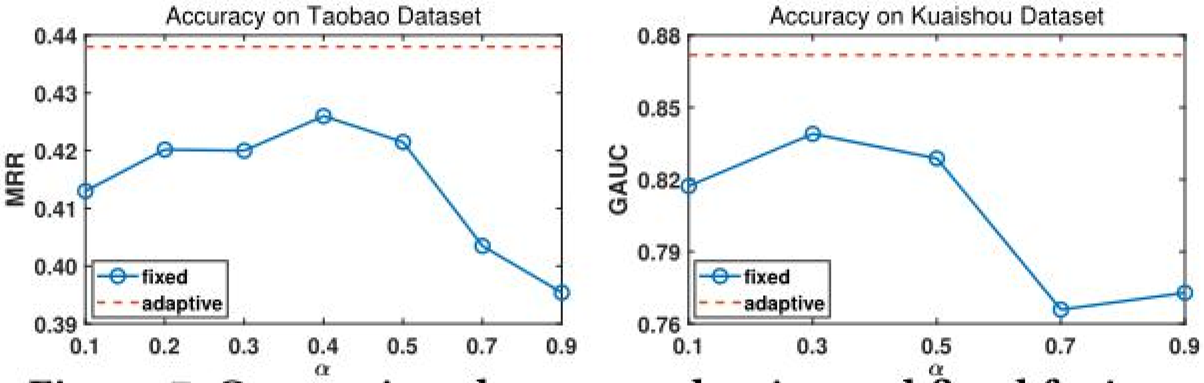


Figure 7: Comparison between adaptive and fixed fusion.

Thanks