Disentangling Long and Short-Term Interests for Recommendation

Yu Zheng, Chen Gao, Jianxin Chang, Yanan Niu, Yang Song, Depeng Jin, Yong Li

Tsinghua University, Kuaishou Technology

ABSTRACT

Modeling user's long-term and short-term interests is crucial for accurate recommendation. 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. In this paper, to address it, we propose a Contrastive learning framework to disentangle Long and Short-term interests for Recommendation (CLSR) with self-supervision. Specifically, we first propose two separate encoders to independently capture user interests of different time scales. We then extract long-term and short-term interests proxies from the interaction sequences, which serve as pseudo labels for user interests. Then pairwise contrastive tasks are designed to supervise the similarity between interest representations and their corresponding interest proxies. Finally, since the importance of long-term and short-term interests is dynamically changing, we propose to adaptively aggregate them through an attention based network for prediction. We conduct experiments on two large-scale real-world datasets for e-commerce and short-video recommendation. Empirical results show that our CLSR consistently outperforms all state-of-the-art models with significant improvements: GAUC is improved by over 0.01, and NDCG is improved by over 4%. Further counterfactual evaluations demonstrate that stronger disentanglement of long and short-term interests is successfully achieved by CLSR. The code and data are available at https://github.com/tsinghua-fib-lab/CLSR.

PROBLEM FORMULATION

Notations. Let M denote the number of users, and $\{x^u\}_{u=1}^{M}$ denote the interaction sequences for all users. Each sequence $x^u = [x_1^u, x_2^u, ..., x_{T_u}^u]$ 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.

Since a user's interaction history x^u reflects both long and short-term interests, the recommender system will first learn LS-term interests from x^u, and then predict future interactions based on the two aspects. We then can formulate the problem of learning LS-term interests for recommendation as follows:

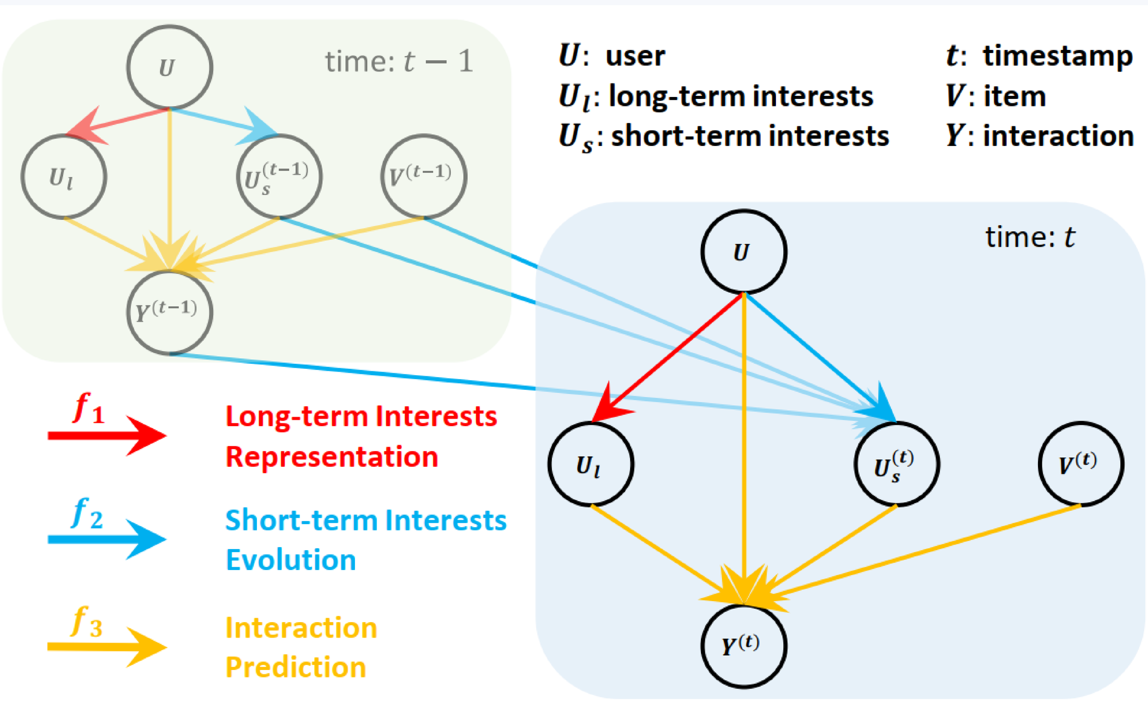
Input: The historical interaction sequences for all users $\{x^u\}_{u=1}^{M}$.

Output: A predictive model that estimates the probability of whether a user will click an item, considering both LS-term interests.

METHOD

User Interests Modeling

- Long-term Interests Representation
- Short-term Interests Evolution
- Interaction Prediction



Self-supervised Implementation

- Generating Query Vectors for LS-term Interests

$$q_l^u = \mathrm{Embed}(u),\quad q_s^{u,t} = \mathrm{GRU}(\{x_1^u, \cdots, x_t^u\}),$$

$$u_l^t = \phi(q_l^u, \{x_1^u, \cdots, x_t^u\}),\quad u_s^t = \psi(q_s^{u,t}, \{x_1^u, \cdots, x_t^u\}),$$

- Interests Encoder

$$v_j = W_l E(x_j^u),$$
$$\alpha_j = \tau_l(v_j \| q_l^u \| (v_j - q_l^u) \| (v_j \cdot q_l^u)),$$
$$a_j = \frac{exp(\alpha_j)}{\sum_{i=1}^{t} exp(\alpha_i)},$$

$$\{o_1^u, ..., o_t^u\} = \rho(\{E(x_1^u), ..., E(x_t^u)\}),\quad v_j = W_s o_j^u,$$

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

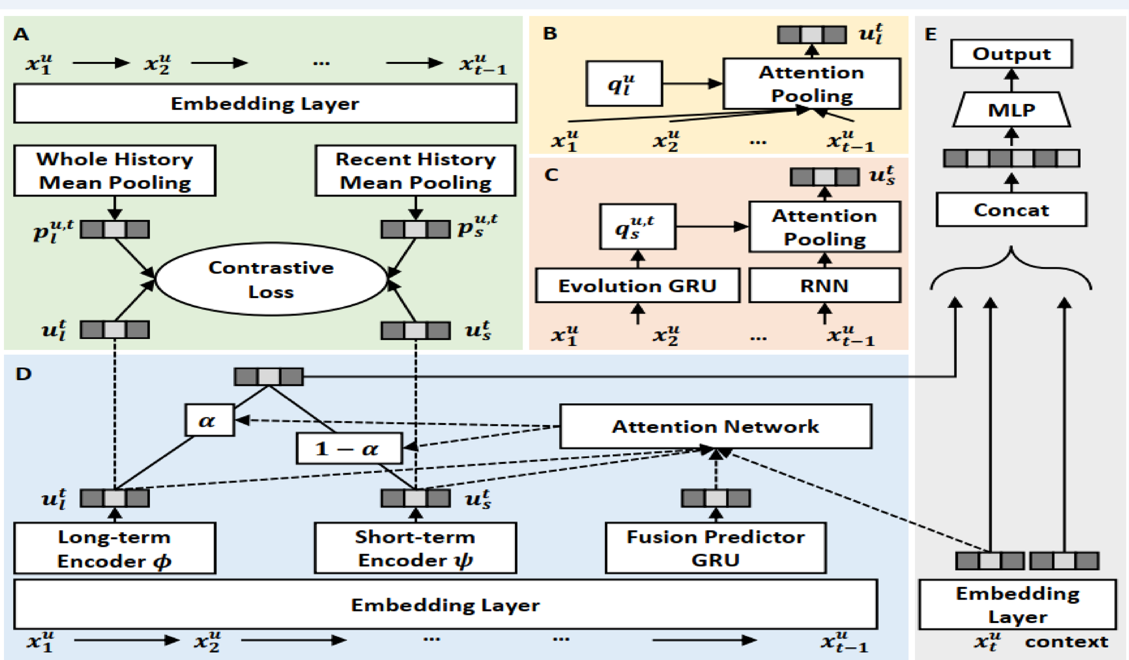
- Self-supervised Disentanglement

$$p_l^{u,t} = \mathrm{MEAN}(\{x_1^u, \cdots, x_t^u\}) = \frac{1}{t}\sum_{j=1}^{t} E(x_j^u),$$
$$p_s^{u,t} = \mathrm{MEAN}(\{x_{t-k+1}^u, \cdots, x_t^u\}) = \frac{1}{k}\sum_{j=1}^{k} E(x_{t-j+1}^u),$$

$$sim(u_l^t, p_l^{u,t}) > sim(u_l^t, p_s^{u,t}),$$
$$sim(p_l^{u,t}, u_l^t) > sim(p_l^{u,t}, u_s^t),$$
$$sim(u_s^t, p_s^{u,t}) > sim(u_s^t, p_l^{u,t}),$$
$$sim(p_s^{u,t}, u_s^t) > sim(p_s^{u,t}, u_l^t),$$

- Adaptive Fusion

$$h_t^u = \mathrm{GRU}(\{E(x_1^u), ..., E(x_t^u)\}),$$
$$\alpha = \sigma(\tau_f(h_t^u \| E(x_{t+1}^u) \| u_l^t \| u_s^t),$$
$$u^t = \alpha \cdot u_l^t + (1-\alpha) \cdot u_s^t,$$

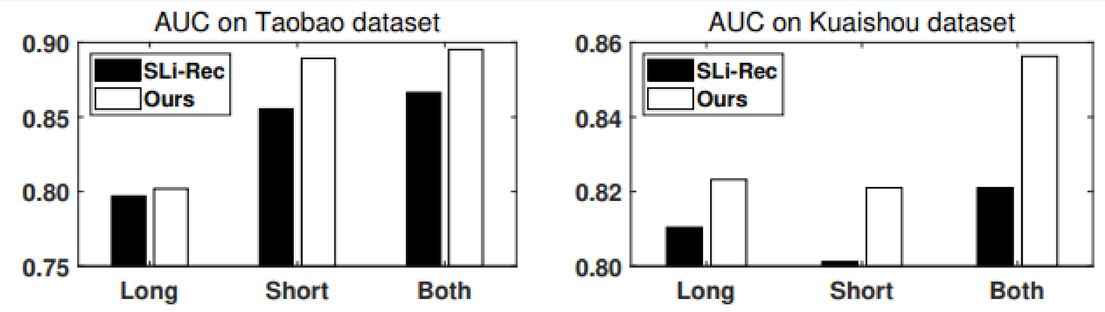


RESULTS

RQ1: How does the proposed framework perform compared with state-of-the-art recommendation models?

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	0.3422	0.8156	0.8298	0.9166	0.9384
	DIEN	0.8477	0.8745	0.4011	0.3404	0.7037	0.7800	0.9030	0.9284
	SASRec	0.8598	0.8635	0.3915	0.3340	0.8199	0.8293	0.9161	0.9380
	SURGE	0.8906	0.8888	0.4228	0.3625	0.8525	0.8610	0.9316	0.9495
LS-term	SLi-Rec	0.8664	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***

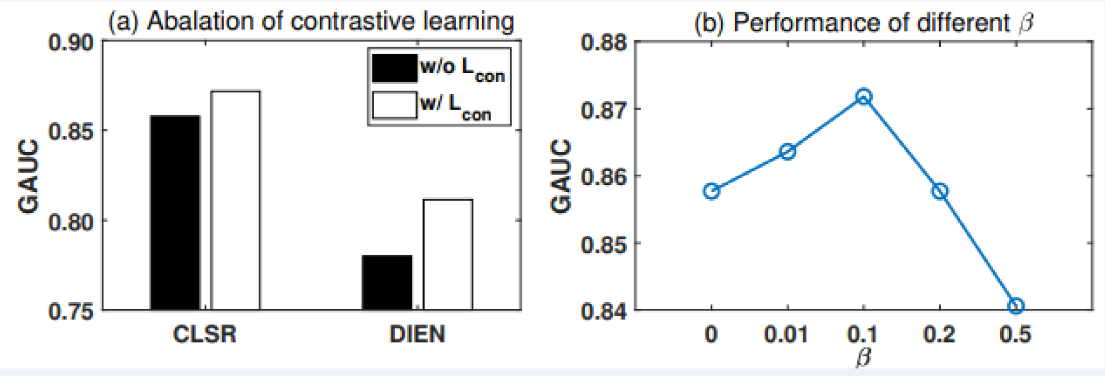
RQ2: Can CLSR achieves stronger disentanglement of LS-term interests against existing unsupervised baselines?

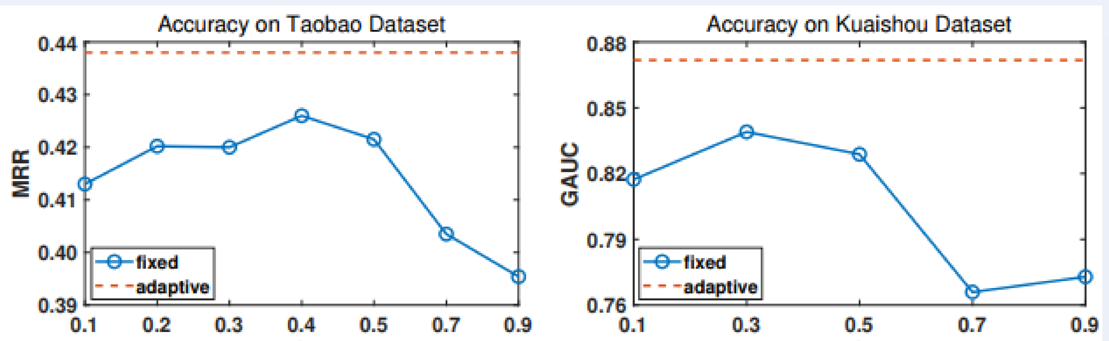


Dataset	Method	Click AUC	Click AVG(α)	Purchase/Like AUC	Purchase/Like AVG(α)
Taobao	SLi-Rec	0.8572	0.4651	0.8288	0.4350 (-6.47%)
	CLSR	0.8885	0.3439	0.8616	0.3568 (**+3.75%**)
Kuaishou	SLi-Rec	0.8153	0.7259	0.7924	0.7543 (+3.91%)
	CLSR	0.8618	0.2528	0.7946	0.2757 (**+9.06%**)

Dataset	Method	Click AUC	Click MRR	Purchase/Like AUC	Purchase/Like 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**

RQ3: What is the effect of different components in CLSR?





CONCLUSIONS

In this paper, we propose to disentangle long and short-term interests for recommendation with a contrastive learning framework, CLSR. Extensive experiments and counterfactual evaluations on two large-scale datasets demonstrate that CLSR consistently outperforms SOTA baselines with significant improvements. More importantly, we empirically show that unsupervised LS-term interests modeling can easily entangle the two aspects and lead to even poorer performance. With the help of self-supervision, CLSR can effectively disentangle LS-term interests and achieve much better performance. As for future work, CLSR can be easily extended since it is a highly general framework, For example, other designs of encoders or proxies can be explored. Deploying the proposed method to industrial systems is another important future work.

REFERENCES

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. Tensorflow: A system for large-scale machine learning. In 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16). 265–283.

[2] Mingxiao An, Fangzhao Wu, Chuhan Wu, Kun Zhang, Zheng Liu, and Xing Xie. 2019. Neural news recommendation with long-and short-term user representations. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 336–345.

[3] Andreas Argyriou, Miguel González-Fierro, and Le Zhang. 2020. Microsoft Recommenders: Best Practices for Production-Ready Recommendation Systems. In Companion Proceedings of the Web Conference 2020. 50–51.

[4] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. 2020. Unsupervised learning of visual features by contrasting cluster assignments. arXiv preprint arXiv:2006.09882 (2020).

ACKNOWLEDGEMENTS

This work is supported in part by The National Key Research and Development Program of China under grant 2020AAA0106000. This work is also supported in part by the National Natural Science Foundation of China under 61972223, U1936217, 61971267 and U20B2060.