Self-Attentive Interest Retrieval Recommender

Min Wu
School of Software Engineering
Xi'an Jiaotong University
Xi'an, China
e-mail:
epouxdefutaba@stu.xjtu.edu.cn

Chen Li*
School of Software Engineering
Xi'an Jiaotong University
Xi'an, China
e-mail: lynnlc@126.com

Lihua Tian
School of Software Engineering
Xi'an Jiaotong University
Xi'an, China
e-mail: lhtian@xjtu.edu.cn

CCET 2022
Code: None

Reported by liang li

Motivation

Details:

- Sequential recommender uses metadata such as item category to capture user's multi-interests. But this method will not reach its expectation due to item's long-tail property. This property will result a large constant of category cannot be effectively activated by the lack of interaction records.
- Another drawback is that may also lead to over-parameterization caused by the massive categories.

Problem Statement

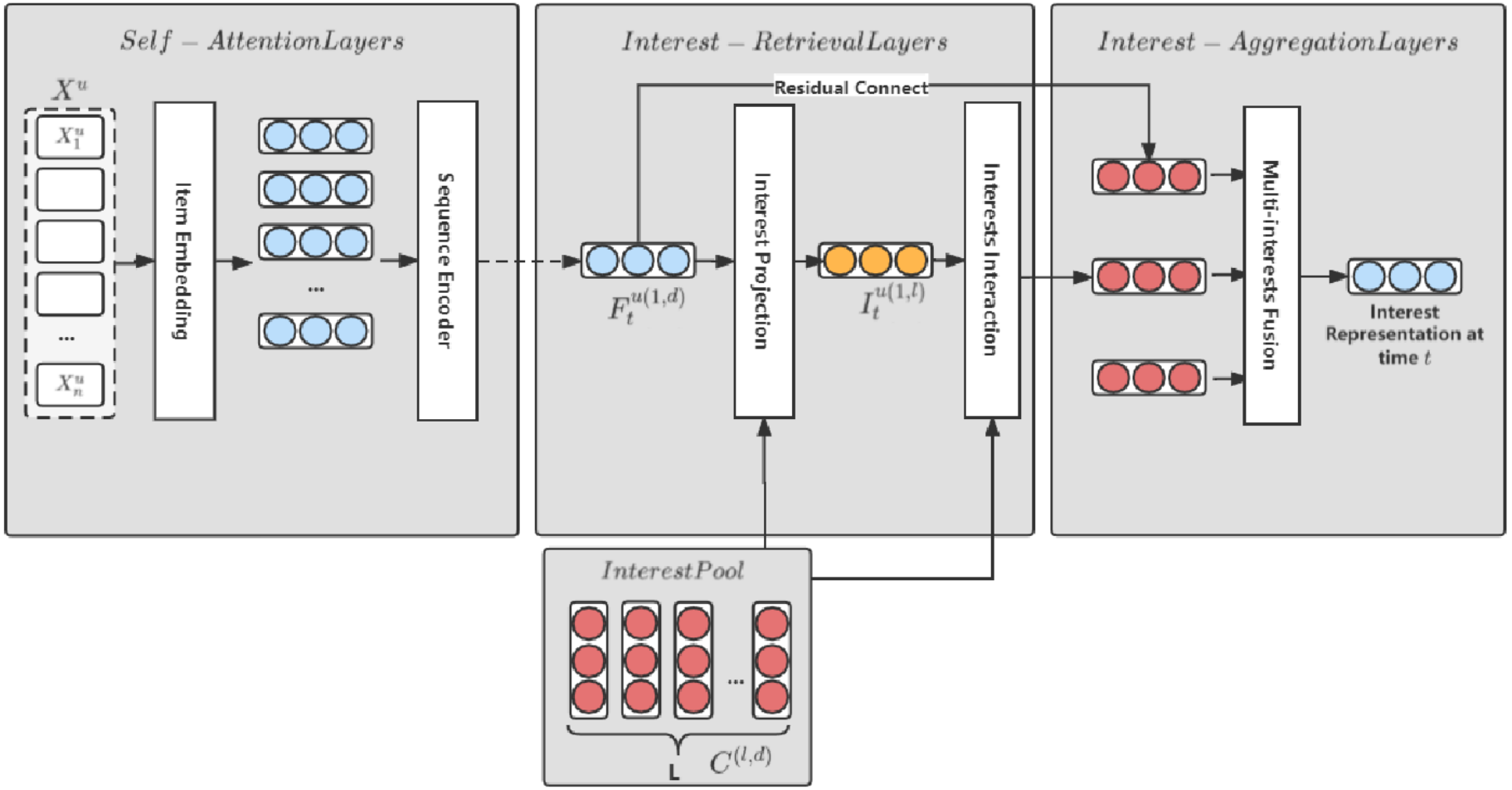

Figure 1. Overall framework of SAIR

$$X^{(u)} = [x_1^{(u)}, x_2^{(u)} \dots x_n^{(u)}]$$

$$H_{id} \in \mathbb{R}^{(MxD)},\ H_{cate} \in \mathbb{R}^{(TxD)},\ H_{pos} \in \mathbb{R}^{(NxD)}.$$

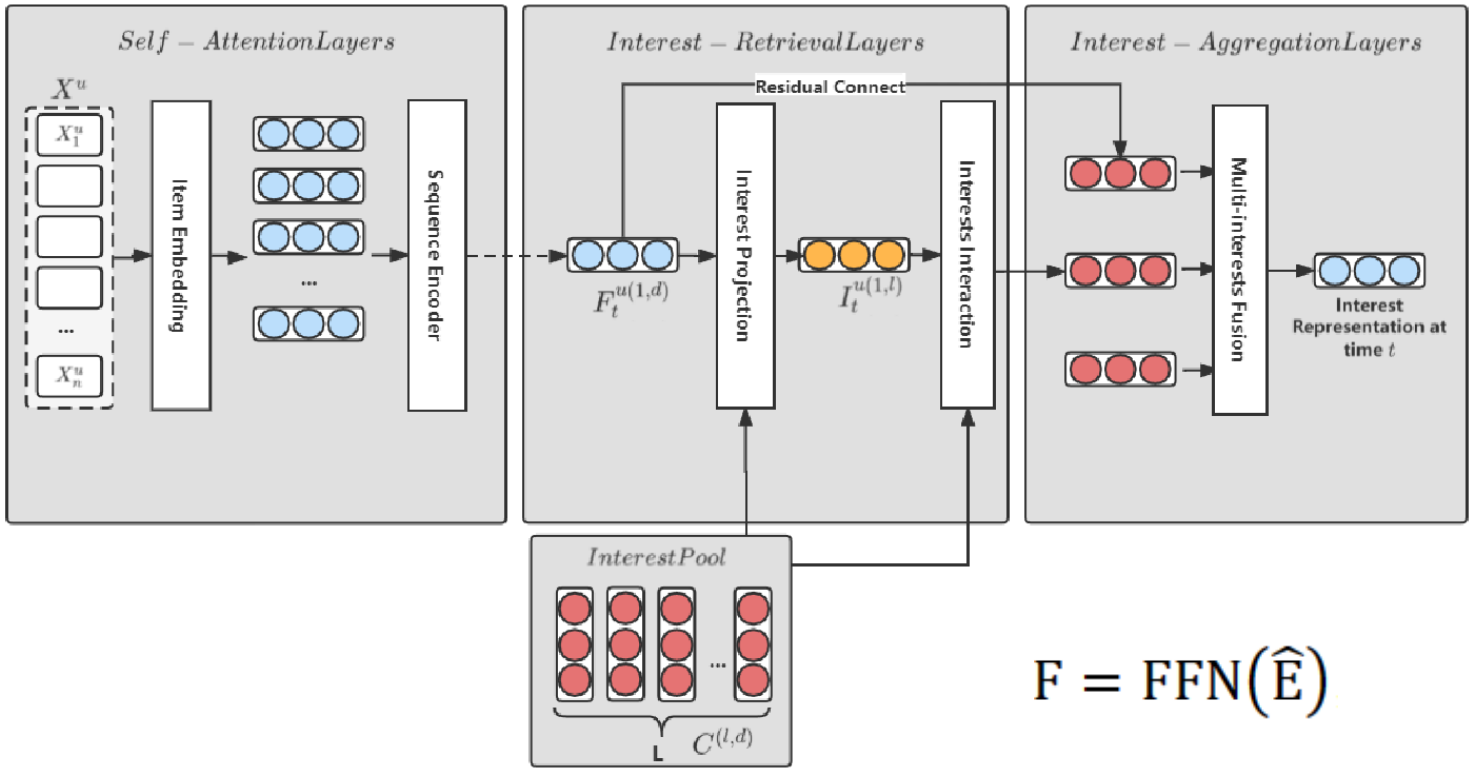

Figure 1. Overall framework of SAIR

$$\mathrm{E}^{(u)} = \mathrm{H}_{\mathrm{id}}(\mathrm{X}^{(u)}) + \mathrm{H}_{\mathrm{cate}}(\mathrm{X}^{(u)}) + \mathrm{H}_{\mathrm{pos}}(\mathrm{X}^{(u)}). \quad (1)$$

$$\hat{\mathrm{E}} = \mathrm{softmax}\left(\frac{\mathrm{QK}}{\sqrt{\mathrm{D}}}\right)\mathrm{V} \quad (2)$$

$$\mathrm{LayerNorm}(\mathrm{Linear}\left(\mathrm{Relu}\left(\mathrm{Dropout}(\mathrm{Linear}(\cdot))\right)\right)). \quad (3)$$

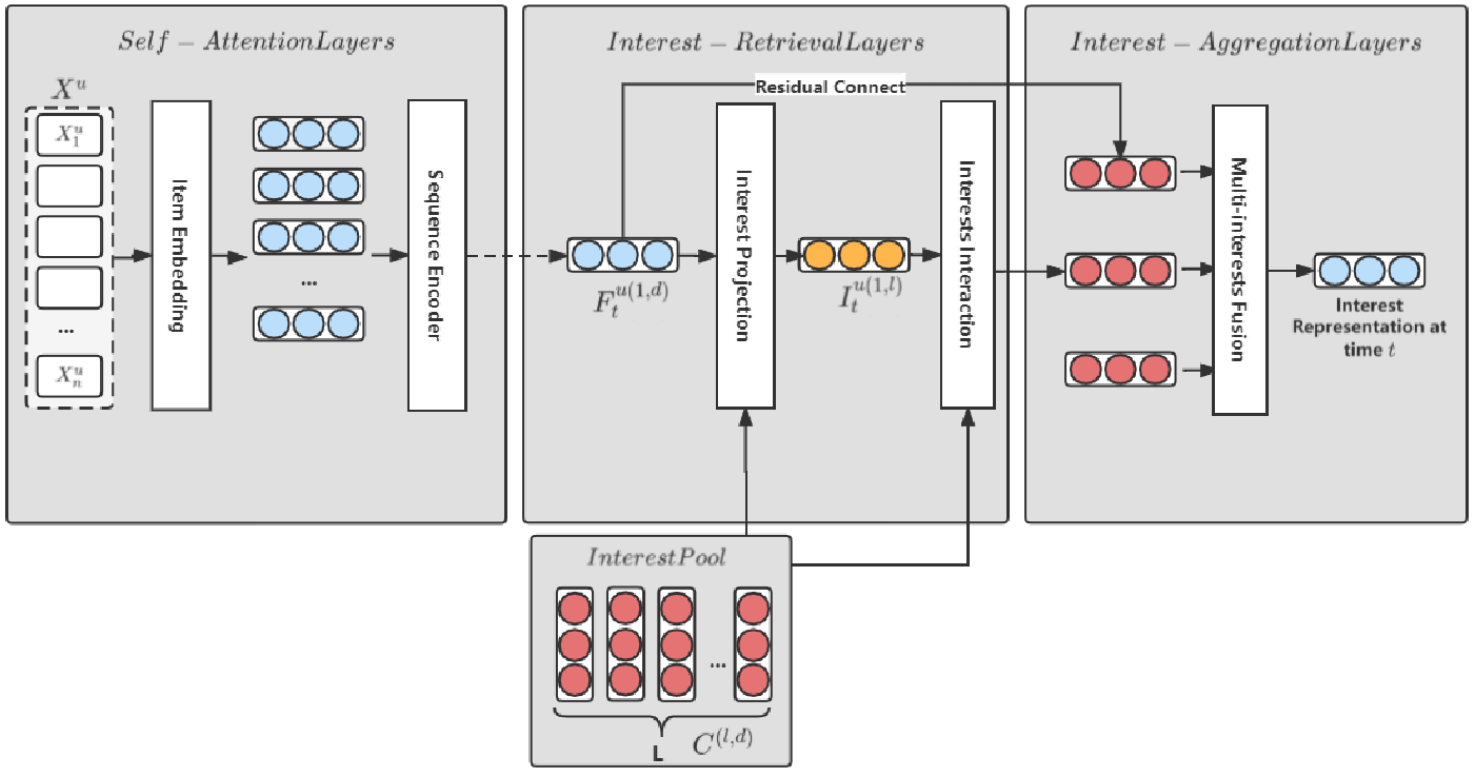

Figure 1. Overall framework of SAIR

$$C_a = F \cdot C^T . \quad (4)$$

$$\widehat{C_a} = \text{Attention}(C_a W^Q, C_a W^K, C_a W^V). \quad (5)$$

$$C_k = \text{top}(\widehat{C_a}, k). \quad (6)$$

$$I_k = \text{Embedding}(C_k, C) \quad (7)$$

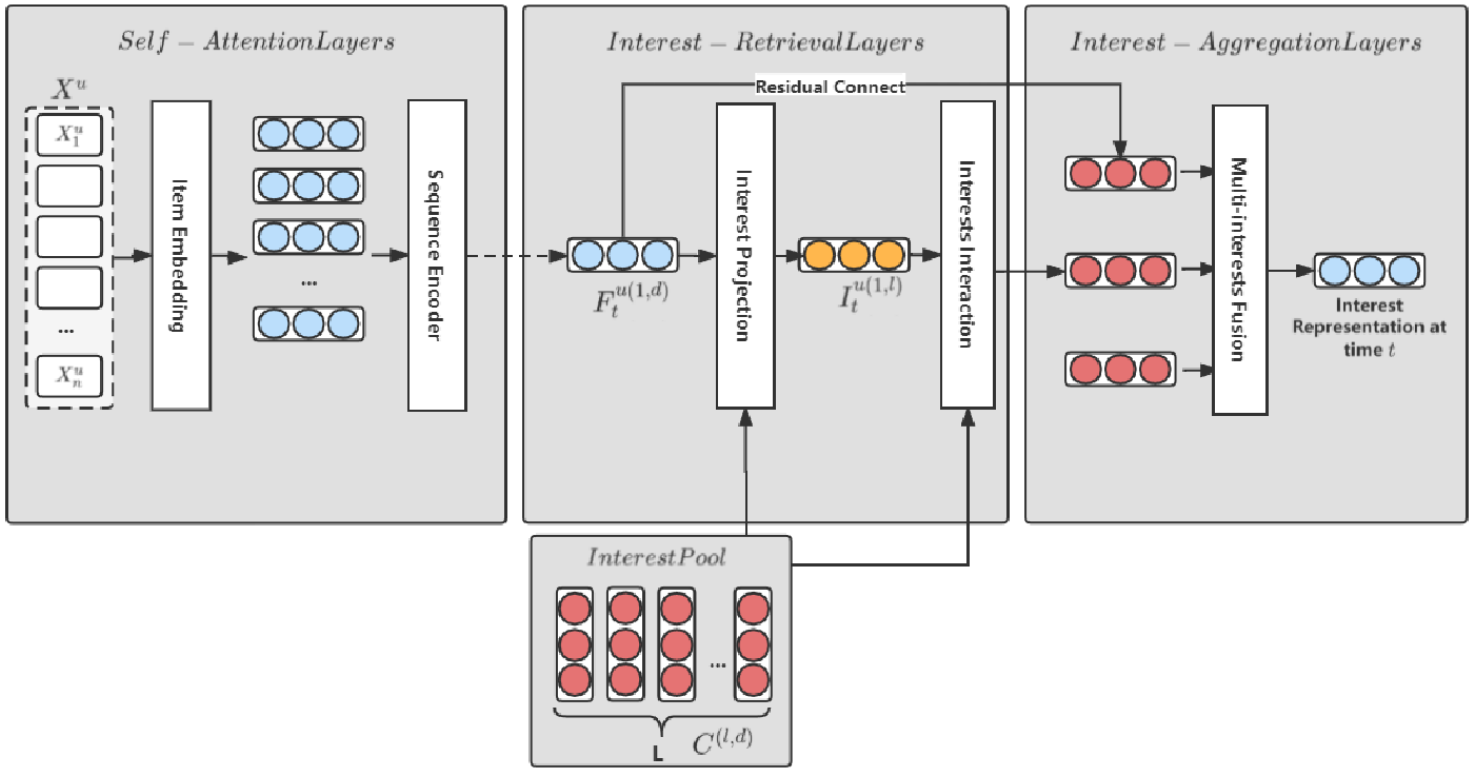

Figure 1. Overall framework of SAIR

$$\hat{I}_k^t = \text{Residual}(I_k^t, F^t) = \left[I_k^{t(1)} + F^t, I_k^{t(2)} + F^t, \ldots, I_k^{t(K)} + F^t\right]. \quad (8)$$

$$e_u^t = \sum_{i=1}^{k} W_i^t \cdot I_a^{t(i)} \quad (10)$$

$$W_i^t = \frac{\exp(C_a^i)}{\sum_j^{C_k} \exp(C_a^j)}, i \in [1, k]. \quad (9)$$

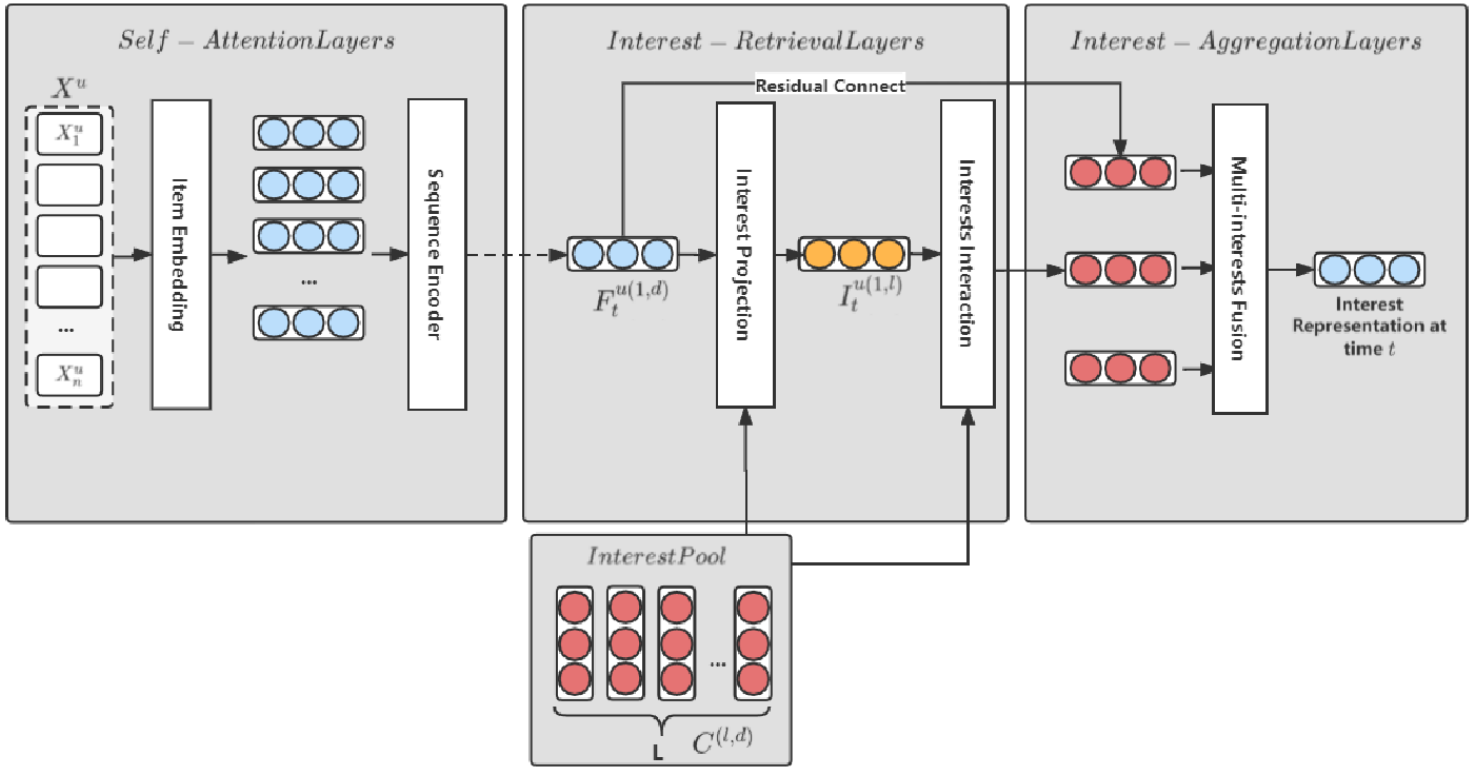

Figure 1. Overall framework of SAIR

$$P(x_u^t | x_u^1, \dots, x_u^{t-1}) = \langle Item_{label}, e_u^t \rangle. \quad (11)$$

$$\mathcal{L}_{ce} = -\sum_{X^u \in X} \sum_{t \in [1,2,\dots,N]} \left[\log(\langle Item_{lab}, e_u^t \rangle) + \log(1 - \langle Item_{neg}, e_u^t \rangle) \right]. \quad (12)$$

$$\mathcal{L}_d = \frac{1}{2} (\|Z\|_F^2 - \|diag(Z)\|_2^2)$$

$$Z = \frac{1}{D} (C - \bar{C})(C - \bar{C})^T$$

$$\mathcal{L} = \mathcal{L}_{ce} + \mathcal{L}_d,$$

Experiments

TABLE I. STATISTICS OF THE DATASETS

Dataset	*ML-1m*	*ML-20m*	*Anime*
#User	6041	138494	73515
#Items	3883	26745	11200
#Actions	1000038	20000263	7813737
#Category	18	21	82
#Avg.Length	163.4	144.41	96.51
#Sparsity	95.58	99.46	99.05

TABLE II. OVERALL COMPARISON OF PERFORMANCE ON NDCG@K AND HR@K(%). THE BEST PERFORMANCE AND SECOND PERFORMANCE ARE DENOTED IN BOLD AND UNDERLINED RESPECTIVELY

Dataset	Metrics		BERT4Rec	SASRec	Caser	LightSANs	SAIR	.Impv
ML-1M	NDCG	@20	0.5996	0.6112	0.5879	0.6145	**0.6213**	1.1%
		@10	0.5779	0.5968	0.5754	0.5945	**0.6088**	2%
		@5	0.5466	0.5582	0.541	0.5621	**0.5732**	1.9%
	HR	@20	0.8664	0.9035	0.8724	0.8828	**0.9107**	0.8%
		@10	0.7811	0.827	0.7906	0.8046	**0.8334**	0.8%
		@5	0.6843	0.7461	0.6889	0.7045	**0.754**	1%
ML-20M	NDCG	@20	0.7539	0.7567	0.7442	0.7893	**0.8176**	3.6%
		@10	0.7433	0.7496	0.7378	0.782	**0.8126**	3.9%
		@5	0.723	0.7298	0.7059	0.7618	**0.798**	4.7%
	HR	@20	0.9485	0.9622	0.943	0.9889	**0.9908**	0.1%
		@10	0.9058	0.9349	0.9254	0.9641	**0.9713**	0.7%
		@5	0.8445	0.8746	0.8816	0.9091	**0.9269**	1.9%
Anime	NDCG	@20	0.5799	0.7803	0.7537	0.8348	**0.8532**	2.2%
		@10	0.5611	0.7753	0.7363	0.8305	**0.8501**	2.4%
		@5	0.5172	0.7601	0.7294	0.8196	**0.8413**	2.6%
	HR	@20	0.9229	0.9695	0.9535	0.9647	**0.9854**	1.6%
		@10	0.8494	0.95	0.9417	0.9481	**0.9734**	2.4%
		@5	0.7152	0.9043	0.8839	0.9146	**0.9466**	3.5%

Experiments

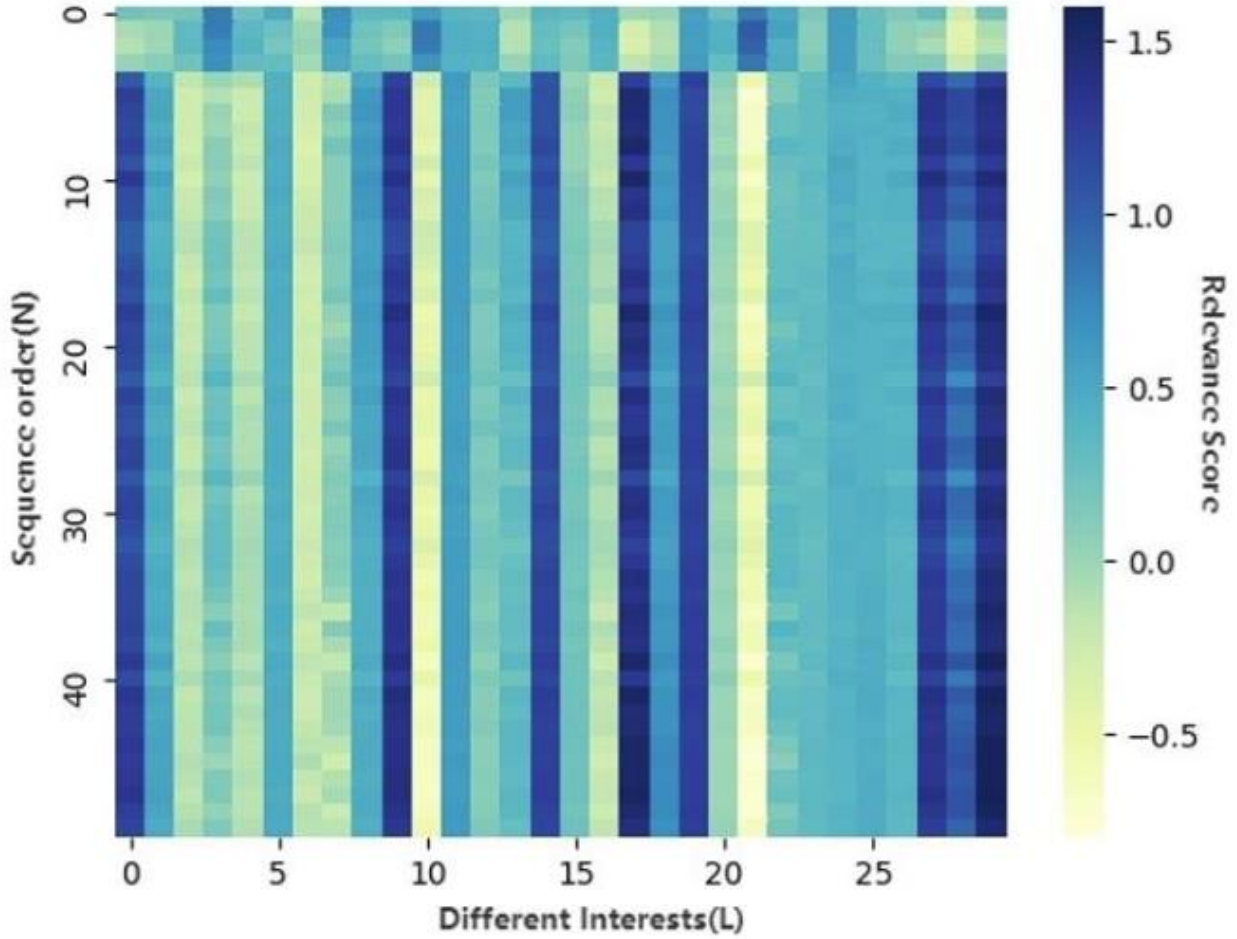

Figure 2. Heat map of distribution matrix.

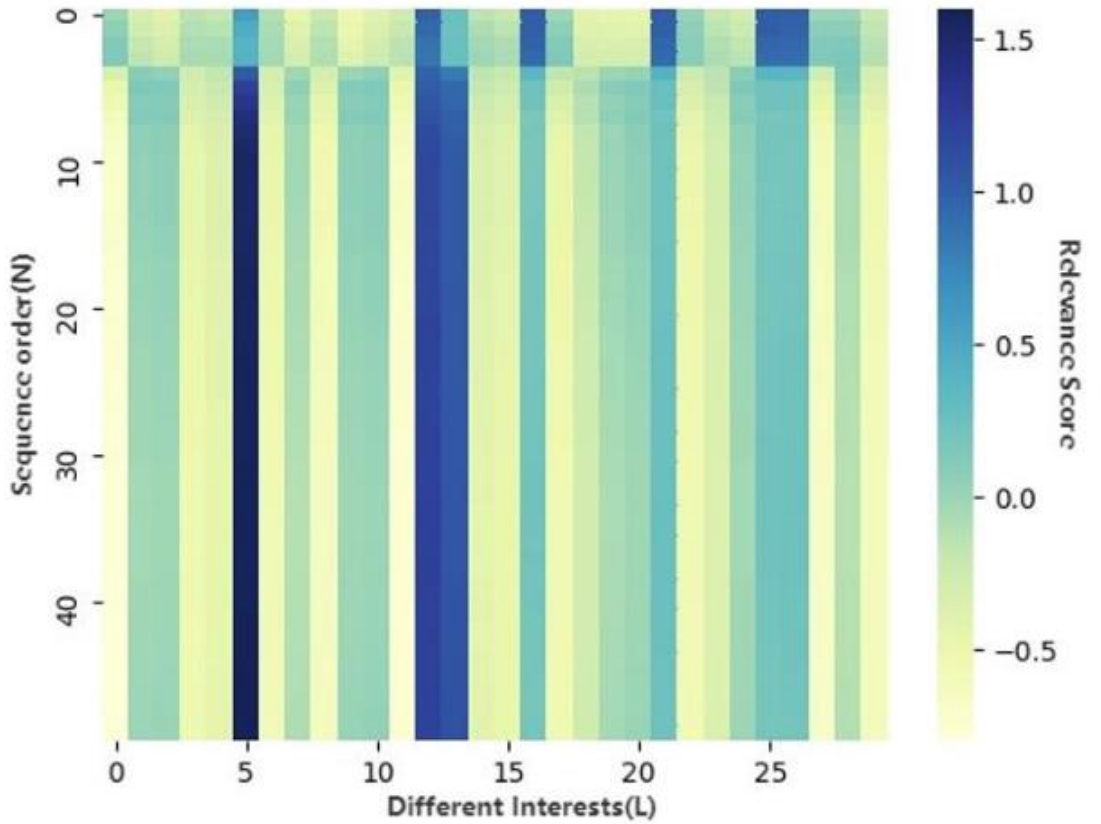

Figure 3. Heat map of distribution processed by interaction layer.

Thanks