Multi-level Contrastive Learning Framework for Sequential Recommendation

Ziyang Wang*
CCIIP Laboratory, Huazhong
University of Science and Technology
Alibaba Group
China
ziyang1997@hust.edu.cn

Huoyu Liu*
Alibaba Group
China
huoyu.lhy@alibaba-inc.com

Wei Wei†
CCIIP Laboratory, Huazhong
University of Science and Technology
Joint Laboratory of HUST and Pingan
Property & Casualty Research (HPL)
China
weiw@hust.edu.cn

Yue Hu
Alibaba Group
China
lingshu.hy@alibaba-inc.com

Xian-Ling Mao
Beijing Institute of Technology
China
maoxl@bit.edu.cn

Shaojian He
Alibaba Group
China
shaojian.he@alibaba-inc.com

Rui Fang
Ping An Property & Casualty
Insurance company of China, Ltd
China
fangrui051@pingan.com.cn

Dangyang Chen
Ping An Property & Casualty
Insurance company of China, Ltd
China
chendangyang273@pingan.com.cn

CIKM 2022

Reported by Nengqiang Xiang

Introduction

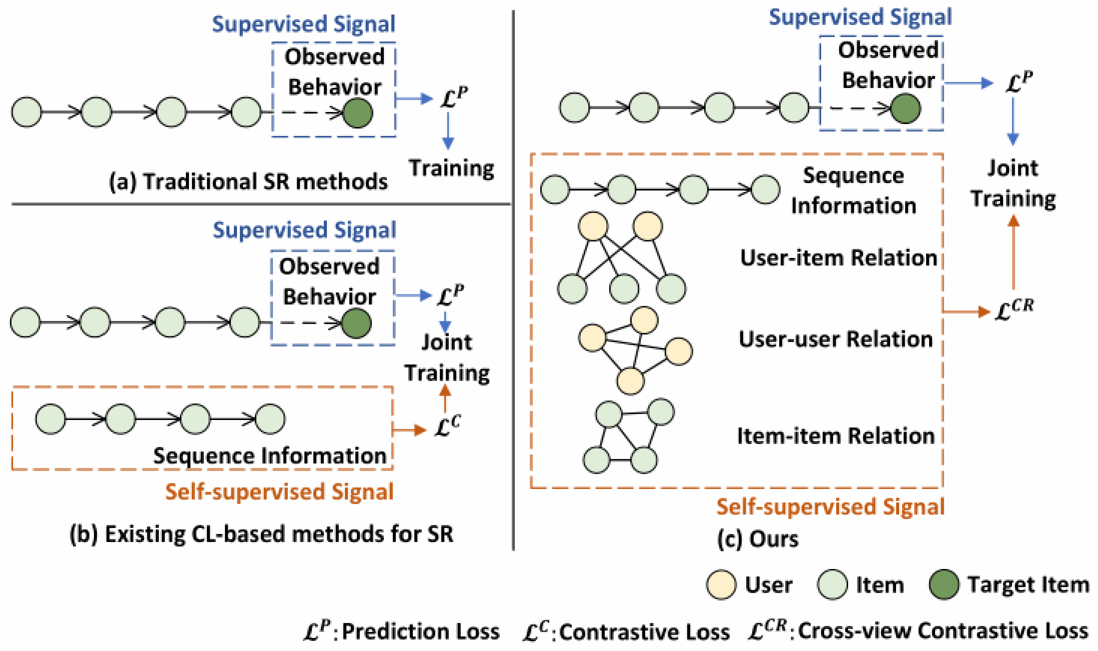


Figure 1: Illustration of training mechanisms of different methods for SR. (a) Traditional methods for SR where the supervised signals are entirely based on the observed user behaviors. (b) Recently contrastive learning-based methods for SR learn the self-supervised signals from the sequence itself. (c) Our proposed method learns rich self-supervised signals by performing cross-view contrastive learning on sequence information, user-item, user-user, and item-item relations.

Each behavior sequence contains a limited number of items, and the self-monitoring information obtained from the sequence is insufficient.

By simply enhancing the data of the behavior sequence, a comparison pair is generated, resulting in low information diversity of the comparison pair.

This paper propose a multi-level contrastive learning framework for sequential recommendation (MCLSR).

Method

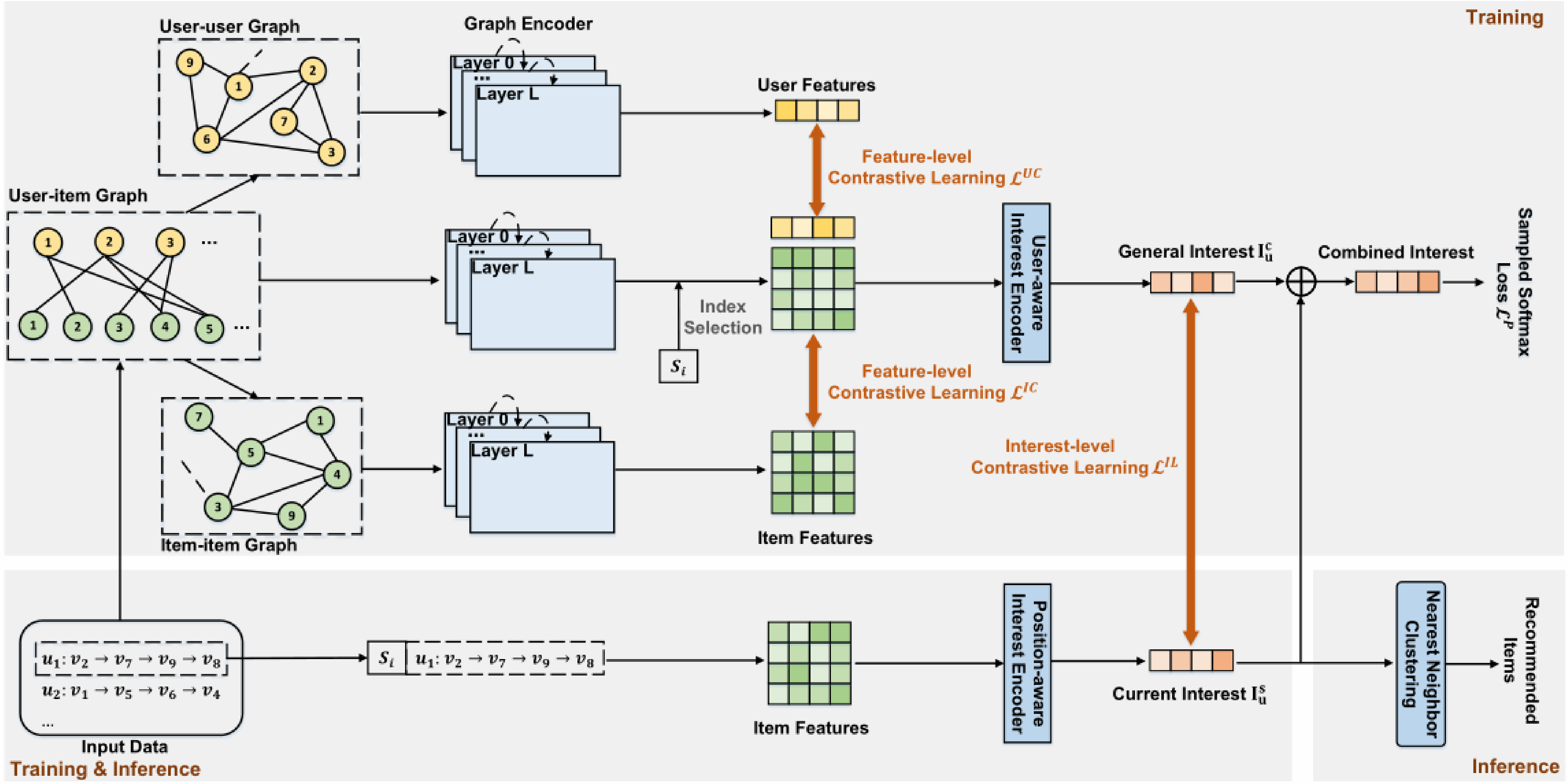
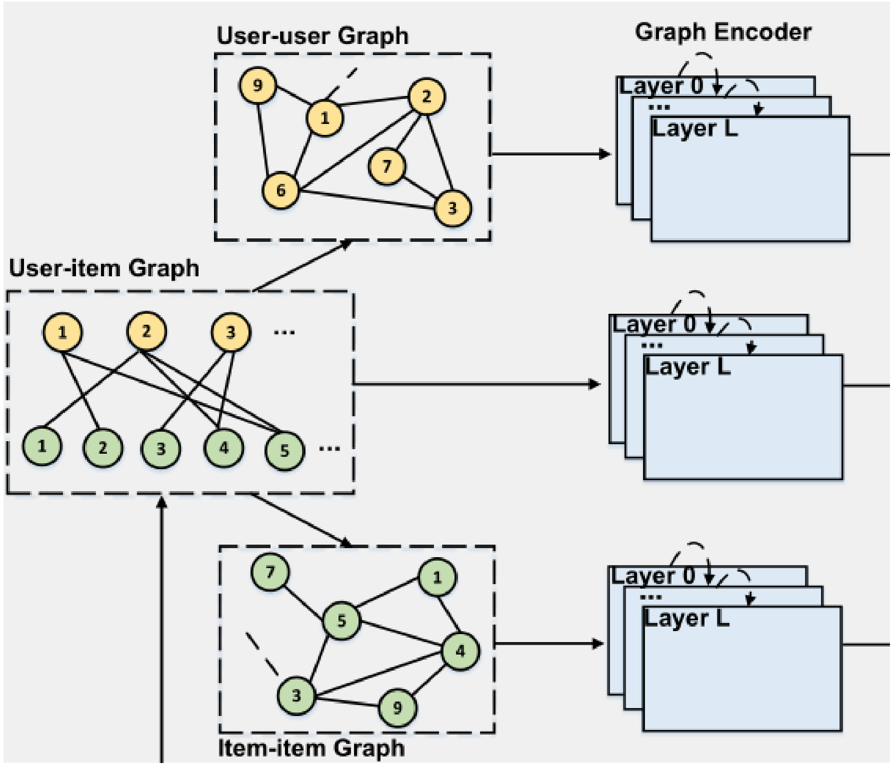


Figure 2: An overview of the proposed framework. ⊕ denotes the element-wise summation.

Method



Problem Formulation:

$u \in \mathcal{U}$: *users set*

$v \in \mathcal{V}$: items *set*

$\mathcal{M}^{uv} \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{V}|}$

$\mathcal{S}^{(u)} = \{v_1^{(u)}, v_2^{(u)}, \ldots v_{|S|}^{(u)}\}$

$\mathrm{H}^u \in \mathcal{R}^{|\mathcal{U}| \times d}$: *user embedding matrix*

$\mathrm{H}^v \in \mathcal{R}^{|\mathcal{V}| \times d}$: *item embedding matrix*

Graph Construction ：

User-item graph:

$$\mathcal{G}^{uv} = (\mathcal{V}^{uv}, \mathcal{E}^{uv})$$

User-user graph:

$$\mathcal{G}^{uv} = (\mathcal{V}^{uu}, \mathcal{E}^{uu})$$

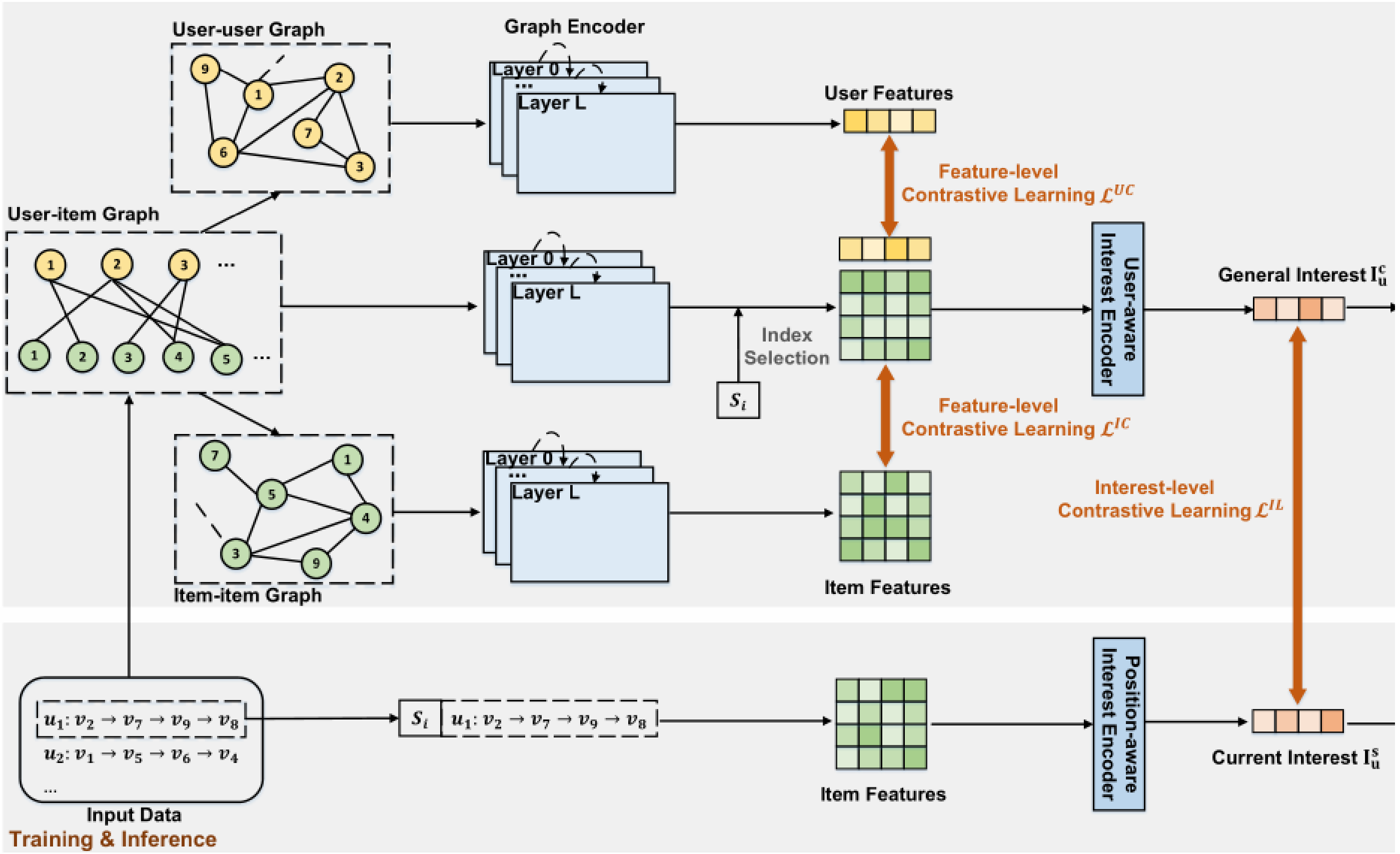
item-item graph ：

$$\mathcal{G}^{uv} = (\mathcal{V}^{vv}, \mathcal{E}^{vv})$$

Graph Encoder Layer:

$$\mathrm{X}^{(l)} = \mathrm{GraphEncoder}(\mathrm{X}, \mathrm{A}) = \mathrm{D}^{-\frac{1}{2}} \mathrm{A} \mathrm{D}^{-\frac{1}{2}} \mathrm{X}^{(l-1)}, \qquad (1)$$

Method



Interest-level Contrastive Learning：

Current interest learning：

$$\mathbf{A}^s = \text{softmax}\left(\mathbf{W}_2 \tanh(\mathbf{W}_1(\mathbf{E}^{u,p})^T)\right), \quad (2)$$

$$\mathbf{E}^{u,p} = [\mathbf{h}_1^v + \mathbf{p}_1, \mathbf{h}_2^v + \mathbf{p}_2, \cdots, \mathbf{h}_n^v + \mathbf{p}_n]$$

$$\mathbf{I}_u^s = \mathbf{A}^s \mathbf{E}^u. \quad (3)$$

$$\mathbf{E}^u = [\mathbf{h}_1^v, \mathbf{h}_2^v, \cdots, \mathbf{h}_n^v]$$

General interest learning:

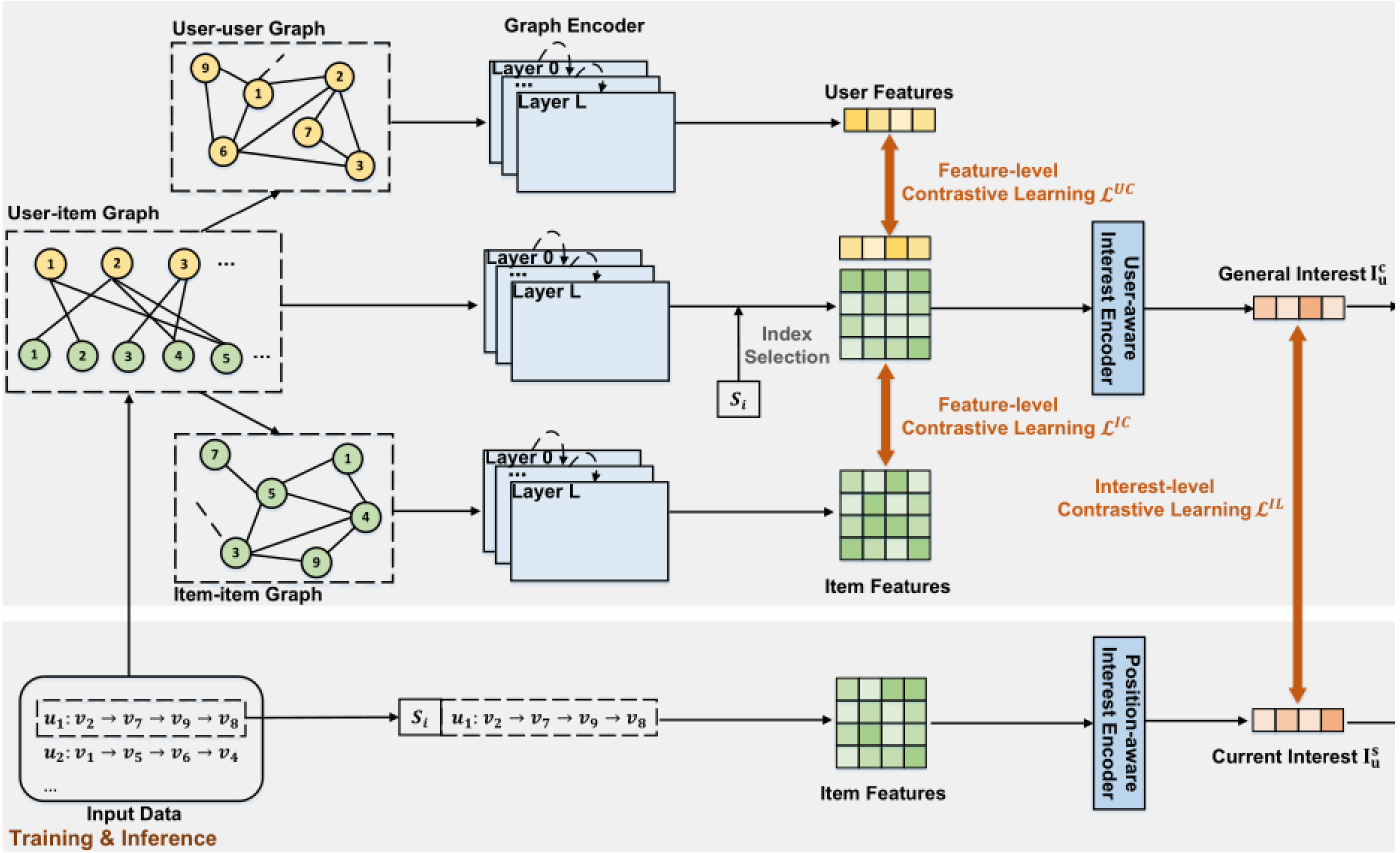
$$\mathbf{H}^{all,uv} = \text{GraphEncoder}^{(l)}(\mathbf{H}^{all}, \mathcal{G}^{uv}), \quad (4)$$

$$\mathbf{H}^{all} = [\mathbf{H}^u || \mathbf{H}^v]$$

$$\mathbf{A}^c = \text{softmax}\left(\tanh\left(\mathbf{W}_3 \mathbf{h}^{u,uv}\right)\left(\mathbf{E}^{u,uv}\right)^T\right), \quad (5)$$

$$\mathbf{I}_u^c = \mathbf{A}^c \mathbf{E}^{u,uv}. \quad (6)$$

Method



Cross-view contrastive learning:

$$\begin{aligned} \mathbf{T}^{I,s} &= \left(\mathbf{W}_2^p \sigma(\mathbf{W}_1^p \mathbf{I}_u^s + \mathbf{b}_1^p) + \mathbf{b}_2^p\right), \\ \mathbf{T}^{I,c} &= \left(\mathbf{W}_2^p \sigma(\mathbf{W}_1^p \mathbf{I}_u^c + \mathbf{b}_1^p) + \mathbf{b}_2^p\right), \end{aligned} \tag{7}$$

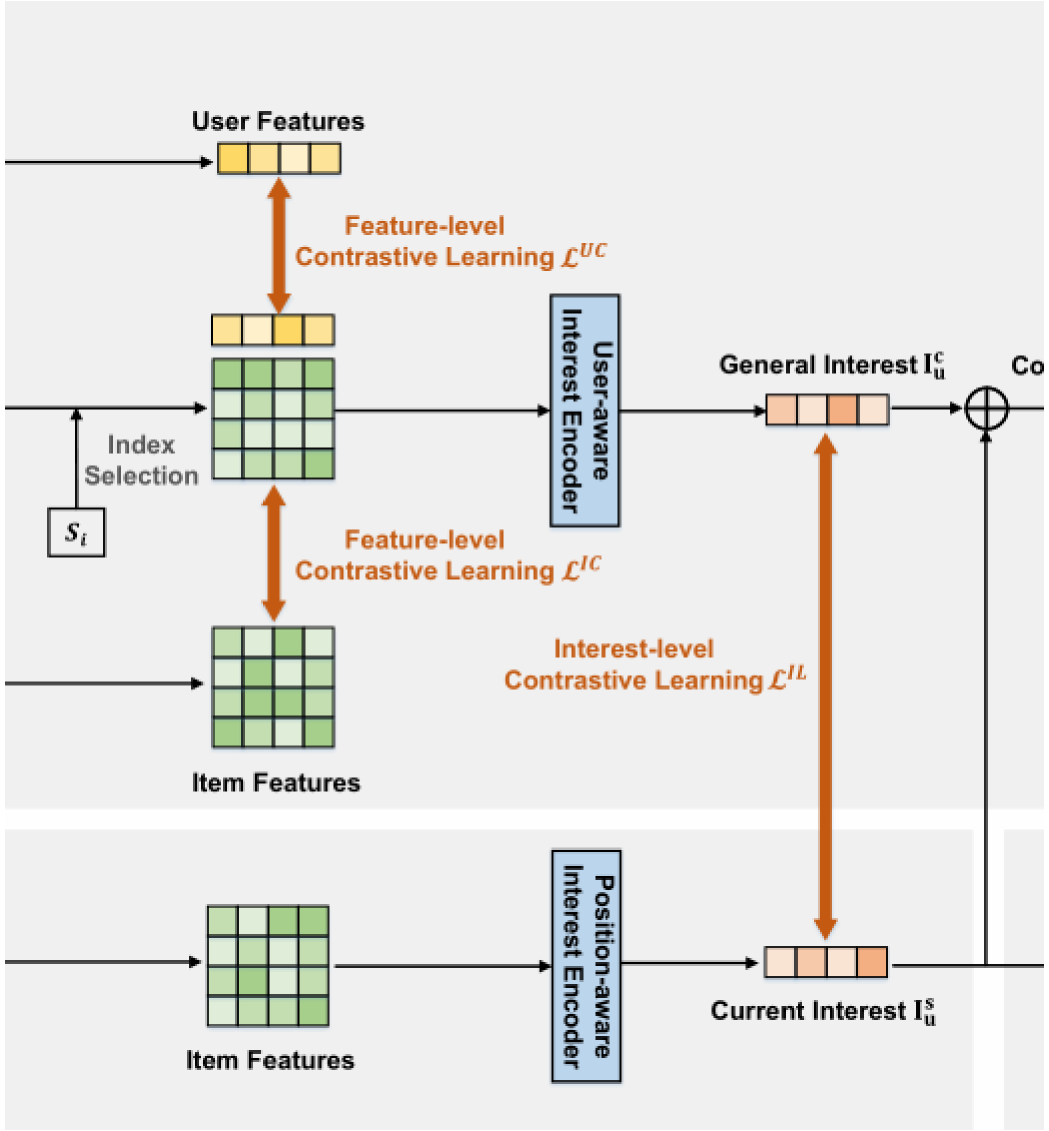
$$\mathcal{L}^{IL} = \sum_{i=1} -\log \frac{\Psi\left(\mathbf{T}_i^{I,s}, \mathbf{T}_i^{I,c}\right)}{\sum_j \Psi\left(\mathbf{T}_i^{I,s}, \mathbf{T}_j^{I,c}\right) + \sum_{j \neq i} \Psi\left(\mathbf{T}_i^{I,s}, \mathbf{T}_j^{I,s}\right)}, \tag{8}$$

Feature-level Contrastive Learning:

Feature learning:

$$\begin{aligned} [\mathbf{H}^{u,uv} || \mathbf{H}^{v,uv}] &= \mathrm{GraphEncoder}^{(l)}([\mathbf{H}^u || \mathbf{H}^v], \mathcal{G}^{uv}), \\ \mathbf{H}^{u,uu} &= \mathrm{GraphEncoder}^{(l)}(\mathbf{H}^u, \mathcal{G}^{uu}), \end{aligned} \tag{9}$$

Method



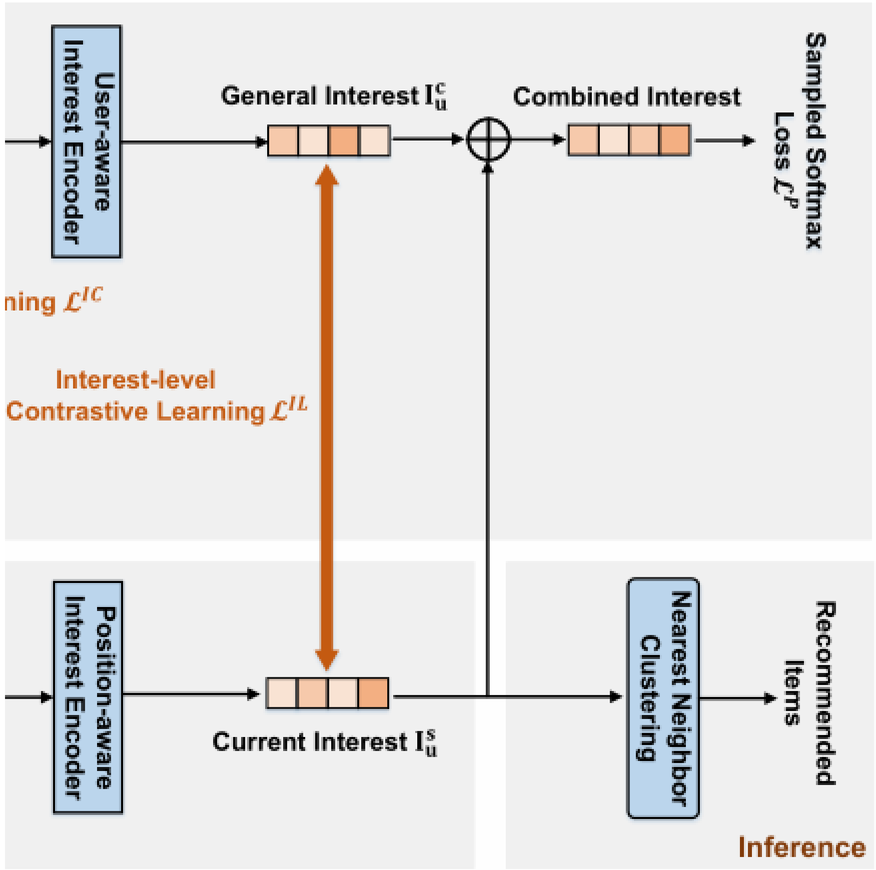
Cross-view contrastive learning:

$$\mathbf{T}^{F,uu} = \mathbf{W}_4^p \sigma(\mathbf{W}_3^p \mathbf{H}^{u,uu} + \mathbf{b}_3^p) + \mathbf{b}_4^p, \quad \mathbf{T}^{F,uv} = \mathbf{W}_4^p \sigma(\mathbf{W}_3^p \mathbf{H}^{u,uv} + \mathbf{b}_3^p) + \mathbf{b}_4^p, \tag{10}$$

$$\mathcal{L}^{UC} = \sum_{i=1} -\log \frac{\Psi\left(\mathbf{T}_i^{F,uv}, \mathbf{T}_i^{F,uu}\right)}{\sum_j \Psi\left(\mathbf{T}_i^{F,uv}, \mathbf{T}_j^{F,uu}\right) + \sum_{j \neq i} \Psi\left(\mathbf{T}_i^{F,uv}, \mathbf{T}_j^{F,uv}\right)}, \tag{11}$$

$$\mathcal{L}^{FL} = \mathcal{L}^{UC} + \mathcal{L}^{IC}. \tag{12}$$

Method



Training and Inference:

Training phase:

$$\mathbf{I}_u^{comb} = \alpha \mathbf{I}_u^s + (1-\alpha)\mathbf{I}_u^c, \tag{13}$$

$$\mathcal{L}^p = \sum_{u \in U} -log \frac{\exp((\mathbf{I}_u^{comb})^T \mathbf{h}_o^v)}{\sum_{k \in Sample(\mathcal{V})} \exp((\mathbf{I}_u^{comb})^T \mathbf{h}_k^v)}. \tag{14}$$

$$\mathcal{J}(\theta) = \mathcal{L}^p + \beta \mathcal{L}^{IL} + \gamma \mathcal{L}^{FL}, \tag{15}$$

Inference phase:

$$R(u, N) = \text{Top-N}_{v \in V}\left((\mathbf{I}_u^s)^T \mathbf{h}_v\right), \tag{16}$$

Experiments

Table 1: Statistics of the used datasets.

Dataset	# user	# item	# interactions	Avg. len.	Sparsity
Books	459,133	313,966	8,898,041	9.7	99.993%
Clothing	39,387	23,034	278,677	6.9	99.969%
Toys	75,258	64,444	1,097,592	9.6	99.977%
Gowalla	65,506	174,606	2,061,264	14.5	99.982%

Experiments

Table 2: Effectiveness comparison between MCLSR and state-of-the-art approaches. † denotes the performance improvement over the best baseline is statistically significant with p-value < 0.01.

Datasets	Metric	POPRec	GRU4Rec	SASRec	ComiRec-SA	GCSAN	S^3-Rec$_{MIP}$	CL4SRec	DuoRec	MCLSR	Improv.
Books	Recall@20	1.368	3.787	6.274	5.489	5.721	6.336	6.544	6.838	**7.469**†	9.2%
	NDCG@20	0.597	1.923	2.825	2.262	2.706	2.964	3.161	3.257	**3.479**†	6.8%
	Hit@20	3.013	8.710	12.765	11.402	11.730	13.052	13.520	14.173	**15.542**†	9.6%
	Recall@50	2.400	6.335	9.349	8.467	8.455	9.684	10.240	10.826	**11.583**†	6.9%
	NDCG@50	0.826	2.600	3.627	3.082	3.434	3.894	4.113	4.308	**4.647**†	7.9%
	Hit@50	5.219	13.597	18.547	17.202	16.865	19.142	20.170	21.366	**23.042**†	7.8%
Clothing	Recall@20	1.200	1.623	2.646	1.678	2.242	2.704	2.863	2.940	**3.138**†	6.7%
	NDCG@20	0.374	0.559	0.854	0.427	0.659	0.873	0.927	1.018	**1.081**†	6.2%
	Hit@20	2.139	2.777	4.188	3.467	3.684	4.343	4.467	4.829	**5.138**†	6.4%
	Recall@50	2.715	2.948	4.505	2.774	3.309	4.522	4.651	4.956	**5.352**†	7.9%
	NDCG@50	0.640	0.778	1.151	0.723	0.829	1.116	1.199	1.356	**1.464**†	8.0%
	Hit@50	4.833	5.085	6.705	5.052	5.812	6.723	7.155	7.785	**8.503**†	9.2%
Toys	Recall@20	0.928	3.214	6.343	5.315	6.593	6.670	6.983	7.841	**8.254**†	10.3%
	NDCG@20	0.510	1.641	2.912	2.114	2.817	3.073	3.072	3.418	**3.726**†	9.0%
	Hit@20	2.496	6.926	12.838	11.075	13.153	13.474	14.079	15.331	**16.661**†	8.7%
	Recall@50	1.844	5.406	10.264	8.962	10.018	10.730	11.300	12.463	**13.328**†	6.9%
	NDCG@50	0.774	2.216	3.899	2.952	3.690	4.072	4.095	4.612	**5.081**†	10.2%
	Hit@50	4.760	11.554	19.837	17.282	19.400	20.363	21.330	23.389	**25.462**†	8.9 %
Gowalla	Recall@20	1.206	5.642	8.581	5.559	7.869	7.823	8.804	8.973	**9.317**†	3.8%
	NDCG@20	1.191	5.536	7.546	3.891	6.819	7.351	7.601	7.618	**7.759**†	1.9%
	Hit@20	5.874	22.450	28.931	19.052	26.315	27.676	29.853	30.075	**31.832**†	5.8%
	Recall@50	2.084	9.623	13.838	9.891	12.793	12.710	14.372	15.195	**15.972**†	5.1%
	NDCG@50	1.678	7.784	10.510	5.725	9.107	9.752	10.630	10.735	**11.012**†	2.6%
	Hit@50	9.716	34.321	42.380	32.041	38.613	39.463	43.659	44.618	**46.217**†	3.5%

Experiments

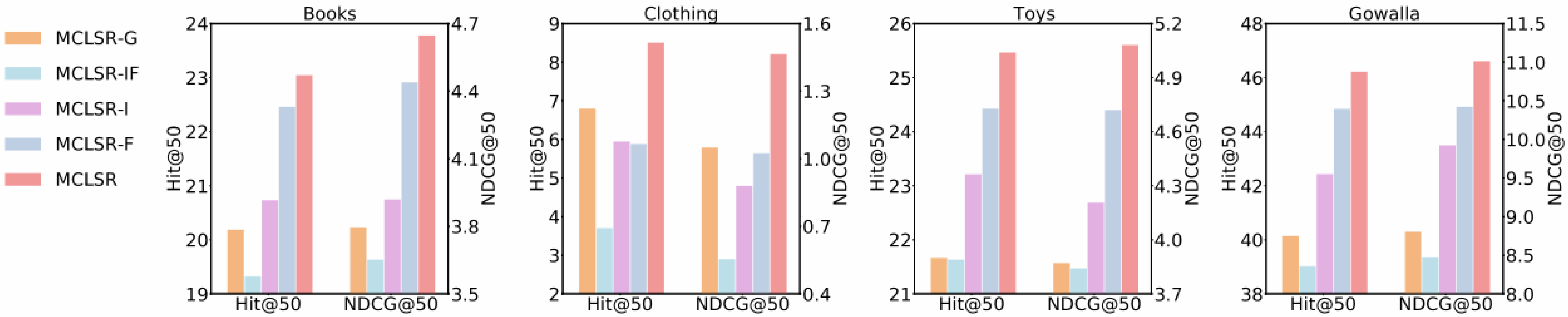
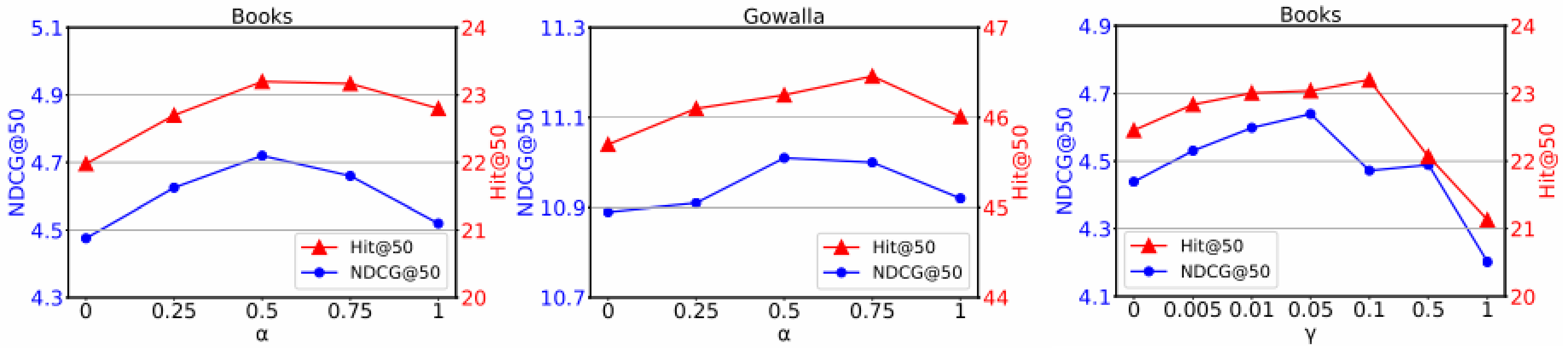


Figure 3: Ablation study on four datasets.

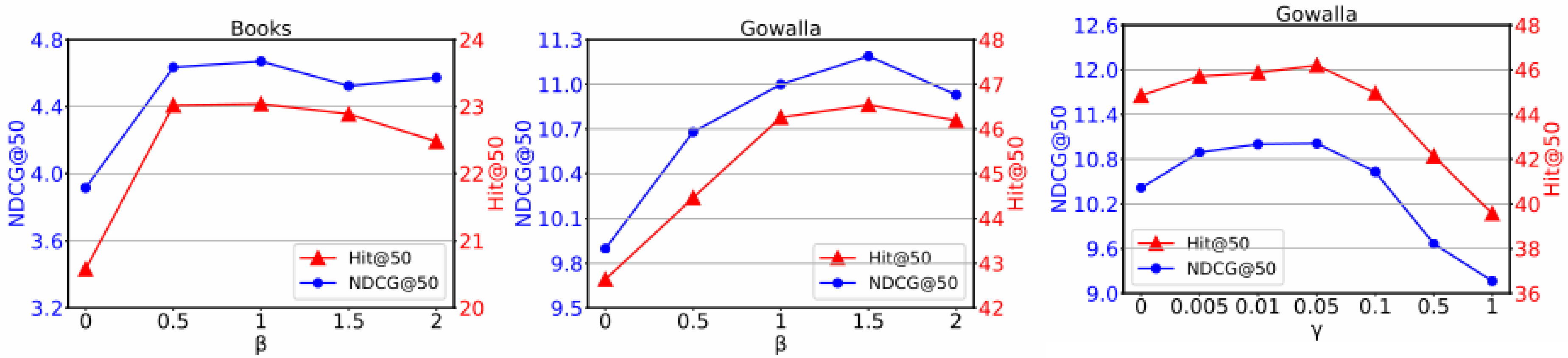
Table 3: The performance of MCLSR with varied depth of GNN layers in terms of Metrics@50.

Depth	Books			Clothing			Toys			Gowalla		
	Recall	NDCG	Hit Rate	Recall	NDCG	Hit Rate	Recall	NDCG	Hit Rate	Recall	NDCG	Hit Rate
$l = 0$	9.881	3.799	20.032	13.025	8.805	40.141	11.052	3.847	21.673	13.265	8.728	40.273
$l = 1$	10.853	4.166	21.782	15.897	10.889	44.543	13.165	4.744	25.069	15.769	10.743	45.620
$l = 2$	**11.583**	**4.647**	**23.042**	**15.972**	**11.012**	**46.217**	13.328	5.081	25.462	**15.972**	**11.012**	**46.217**
$l = 3$	10.085	3.936	20.547	15.021	10.042	43.775	**13.720**	**5.187**	**26.417**	14.948	10.094	43.665

Experiments



(a) Sensitive of parameter α



(b) Sensitive of parameter β

(c) Sensitive of parameter γ

Experiments

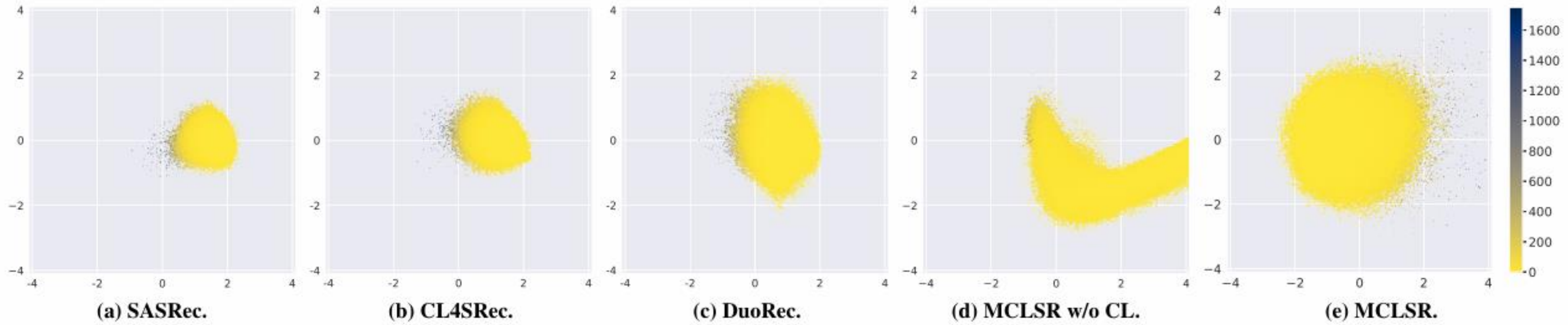


Figure 5: Item embeddings of selected methods on Book dataset.

Thanks