Multi-Intention Oriented Contrastive Learning for Sequential Recommendation

Xuewei Li
College of Intelligence and Computing, Tianjin University
Tianjin, China
lixuewei@tju.edu.cn

Aitong Sun
College of Intelligence and Computing, Tianjin University
Tianjin, China
aitongsun@tju.edu.cn

Mankun Zhao
College of Intelligence and Computing, Tianjin University
Tianjin, China
zmk@tju.edu.cn

Jian Yu
College of Intelligence and Computing, Tianjin University
Tianjin, China
yujian@tju.edu.cn

Kun Zhu
Tianjin International Engineering Institute, Tianjin University
Tianjin, China
2019229048@tju.edu.cn

Di Jin
College of Intelligence and Computing, Tianjin University
Tianjin, China
jindi@tju.edu.cn

WSDM 2023

Reported by Nengqiang Xiang

Introduction

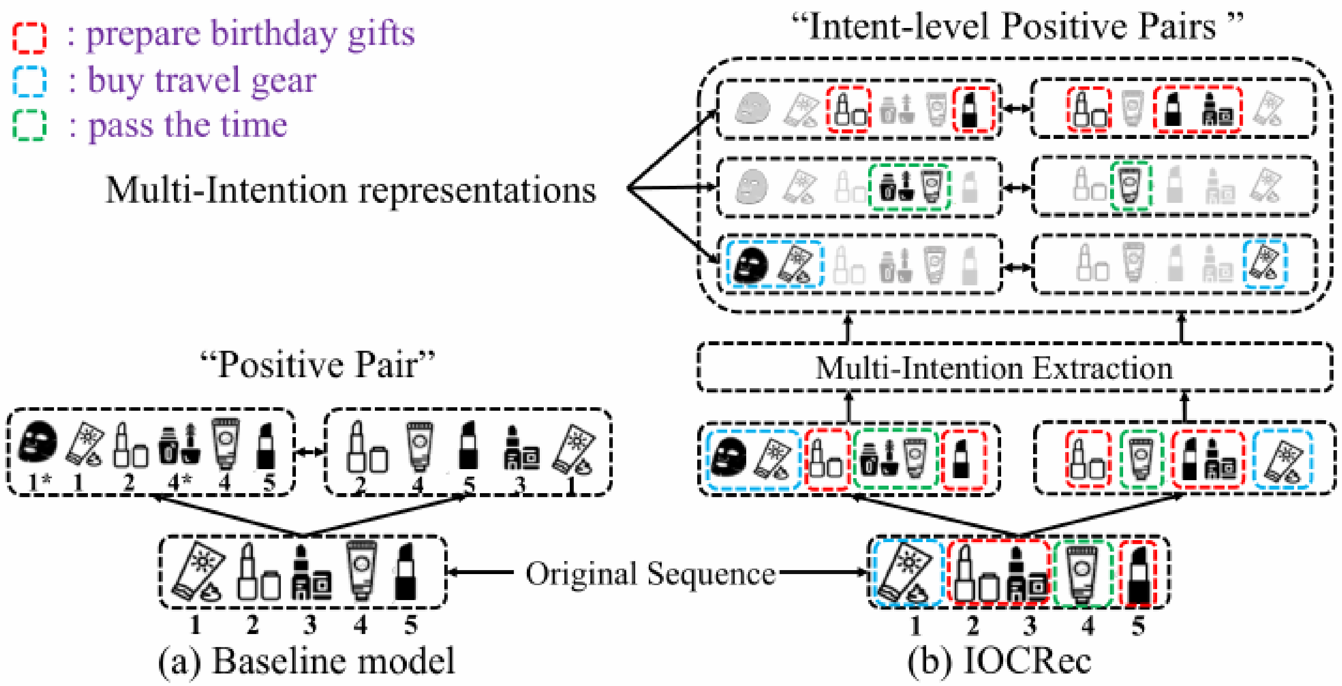


Figure 1: A case of contrastive learning strategies comparing baseline model and our new method IOCRec.

Existing models use data enhancement and contrast learning to solve the problem of data sparsity, but amplify the noise in the original sequence.

Noise interferes with the user's main intention, which makes the two views of contrast learning inconsistent.

In order to solve this problem, the author selects the user's main intention to de-noise and proposes multi-intention oriented comparative learning.

Method

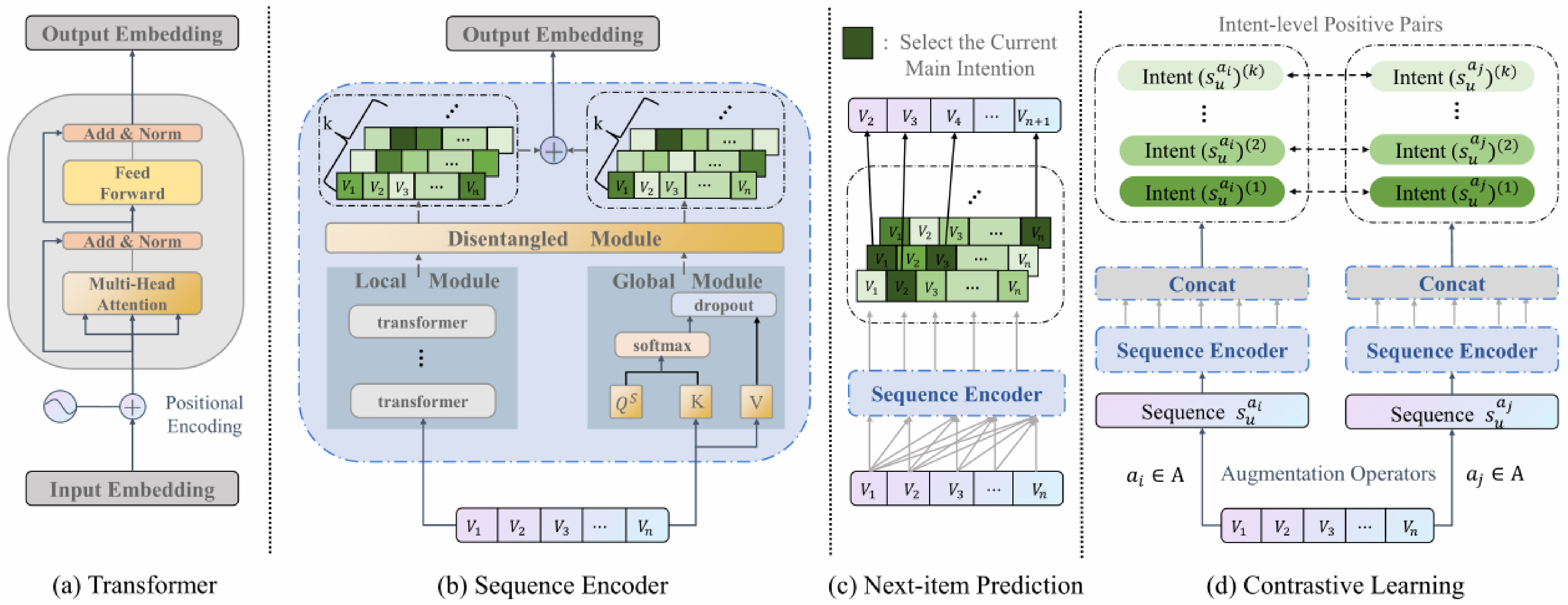


Figure 2: Overall framework. (a) illustrates the structure of Transformer. (b) presents the structure of Sequence Encoder. (c) predicts the next item based on selecting the current main intention. (d) demonstrates CL of a sequence. The augmentation operators set A = {C, M, R, S, I}, the details see Section 2.3, it first augments a sequence as positive pair with random select two augmentation operators a_i,a_j from set A. Then, it encodes the sequence by concatenating embedding outputs from Sequence Encoder. Finally, it maximizes the agreement between intent-level positive pairs.

Method

PROBLEM FORMULATION:

Problem Definition

$$\underset{v_i \in \mathcal{V}}{\arg\max} P\left(v_{|s_u|+1} = v_i \mid s_u\right), \quad (1)$$

Multi-Intention Definition

$$\mathbf{c}_u = [\mathbf{c}_u^{(1)}; \mathbf{c}_u^{(2)}; \ldots; \mathbf{c}_u^{(k)}] \in \mathbb{R}^d$$

Data Augmentation Operators

Crop (C):

$$s_u^{\text{Crop}} = \left[v_c, v_{c+1}, \ldots, v_{c+L_c-1}\right], \quad (2)$$

Mask (M):

$$s_u^{\text{Mask}} = \left[\widehat{v}_1, \widehat{v}_2, \ldots, \widehat{v}_{|s_u|}\right], \quad (3)$$

Reorder (R):

$$s_u^{\text{Reorder}} = \left[v_1, \ldots, \widehat{v}_i, \ldots, \widehat{v}_{i+L_R-1}, \ldots, v_{|s_u|}\right], \quad (4)$$

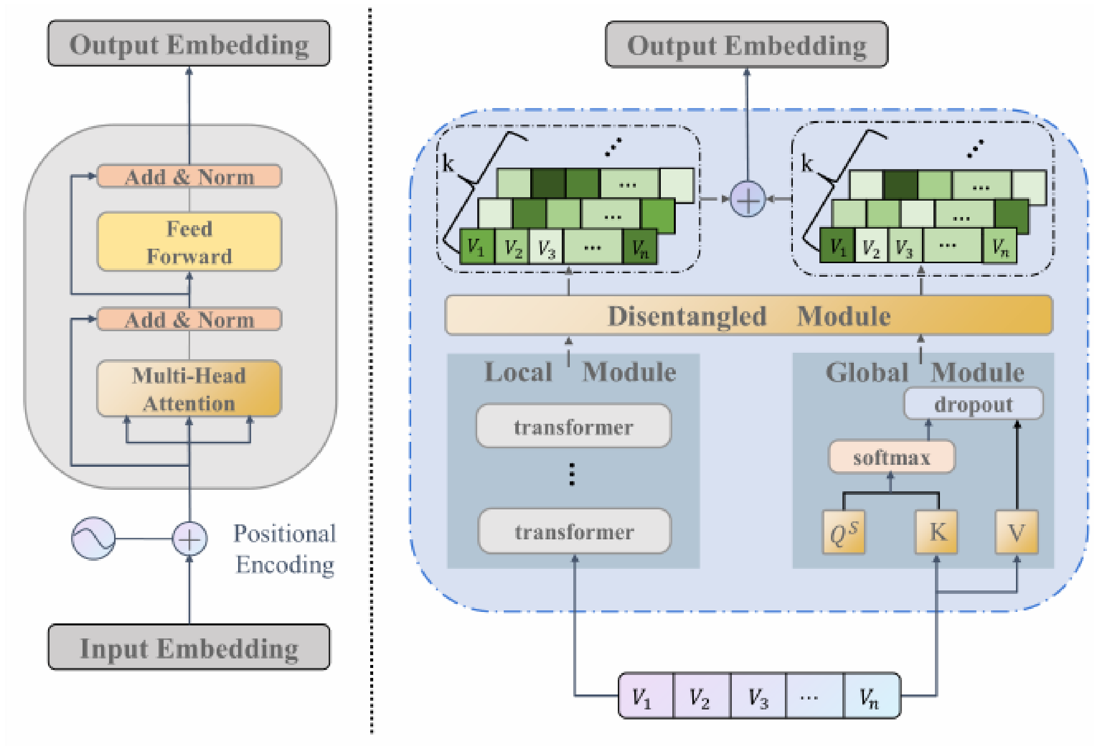
Substitute (S):

$$s_u^{\text{Substitute}} = \left[v_1, v_2, \ldots, \widehat{v}_{idx_i}, \ldots, v_{|s_u|}\right], \quad (5)$$

Insert (I):

$$s_u^{\text{Insert}} = \left[v_1, v_2, \ldots, \widehat{v}_{\text{idx}_i}, v_{\text{idx}_i}, \ldots, v_n\right]. \quad (6)$$

Method



(a) Transformer (b) Sequence Encoder

Local Module

$$\mathbf{P} = [p_1; p_2; \ldots; p_L] \in \mathbb{R}^{L \times d}$$

$$\mathbf{E}_P^{(l)} = Local^{(l)}\left(\mathbf{E}_P^{(l-1)}\right), l \in \{1, 2, \ldots\}, \tag{7}$$

$$\mathrm{MHA}\left(\mathbf{E}_P^{(l)}\right) = \mathrm{concat}\left(\mathrm{head}_1; \cdots; \mathrm{head}_h\right)\mathbf{W}^O,$$

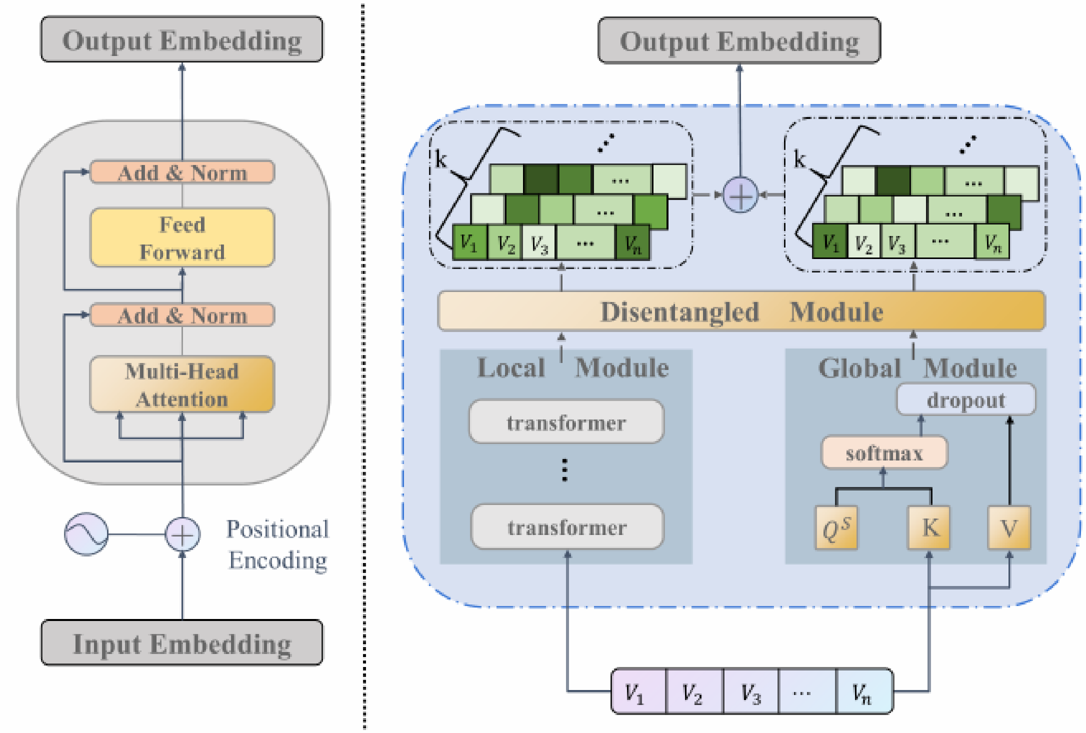
$$\mathrm{head}_i\left(\mathbf{E}_P^{(l)}\right) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d/h}}\right)\mathbf{V}, \tag{8}$$

$$Local\left(\mathbf{E}_P^{(l)}\right) = \left[\mathrm{FFN}\left(\mathbf{E}_{P1}^{(l)}\right)^\top; \cdots; \mathrm{FFN}\left(\mathbf{E}_{Pn}^{(l)}\right)^\top\right].$$

Global Module

$$Global(\mathbf{E}) = \mathrm{Dropout}(\mathrm{softmax}\left(\mathbf{Q}^S\left(\mathbf{E}\mathbf{W}_K'\right)^T\right)\mathbf{E}\mathbf{W}_V'). \tag{9}$$

Method



(a) Transformer (b) Sequence Encoder

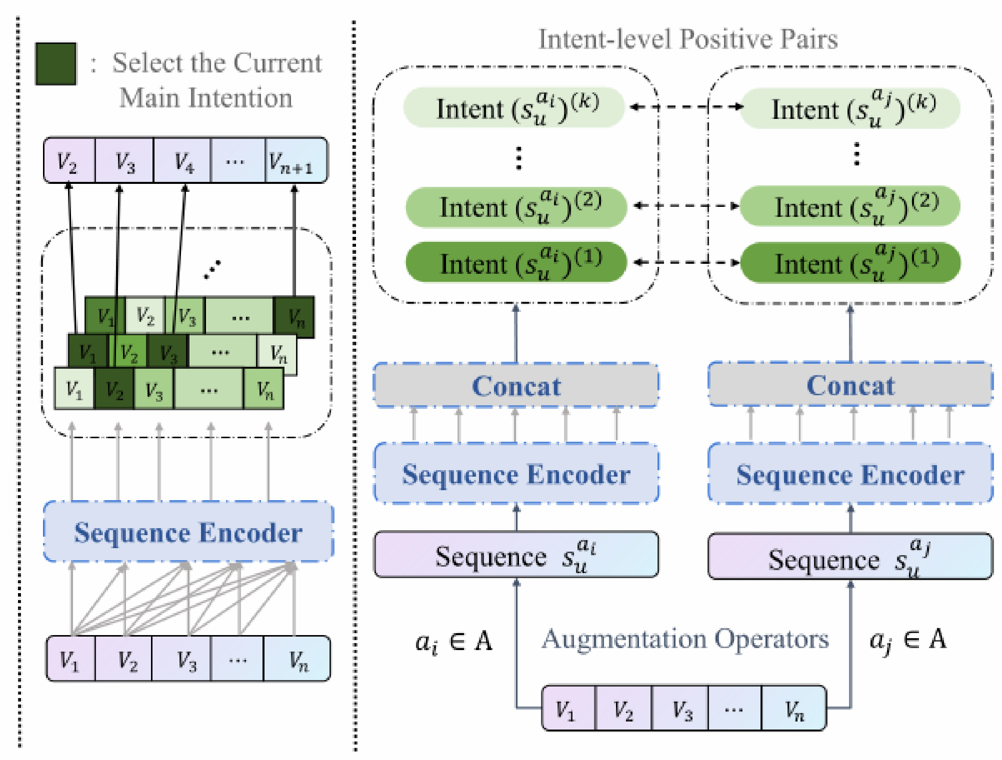
Disentangled Module

$$p_{k|i} = \frac{\exp\left(\frac{1}{\sqrt{d}}\mathrm{LN}_1\left(\mathbf{e}_u^{(i)}\right)\cdot \mathrm{LN}_2\left(\mathbf{c}_u^{(k)}\right)\right)}{\sum_{k'=1}^{K}\exp\left(\frac{1}{\sqrt{d}}\mathrm{LN}_1\left(\mathbf{e}_u^{(i)}\right)\cdot \mathrm{LN}_2\left(\mathbf{c}_u^{(k')}\right)\right)}, \tag{10}$$

$$\begin{aligned} p_i &= \frac{\exp\left(\frac{1}{\sqrt{d}}\mathrm{key}_i \cdot \mathrm{query}\right)}{\sum_{i'=1}^{L}\exp\left(\frac{1}{\sqrt{d}}\mathrm{key}_{i'} \cdot \mathrm{query}\right)}, \\ \mathrm{query} &= \mathrm{LN}_3\left(\varphi_t + \mathrm{e}_u^{(t)} + \rho\right), \\ \widehat{\mathrm{key}}_i &= \mathrm{LN}_4\left(\varphi_i + \mathrm{e}_u^{(i)}\right), \\ \mathrm{key}_i &= \widehat{\mathrm{key}}_i + \mathrm{ReLU}\left(\mathbf{W}^\top \widehat{\mathrm{key}}_i\right), \end{aligned} \tag{11}$$

$$\begin{aligned} LI(s_u^{(i)})^{(k)} &= \mathrm{LN}_5\left(p_{k|i}\cdot p_i \cdot Local(s_u^{(i)})\right), \\ GI(s_u^{(i)})^{(k)} &= \mathrm{LN}_5\left(p_{k|i}\cdot p_i \cdot Global(s_u^{(i)})\right), \\ Final\ Intentions(s_u^{(i)})^{(k)} &= LI(s_u^{(i)})^{(k)} + GI(s_u^{(i)})^{(k)}. \end{aligned} \tag{12}$$

Method



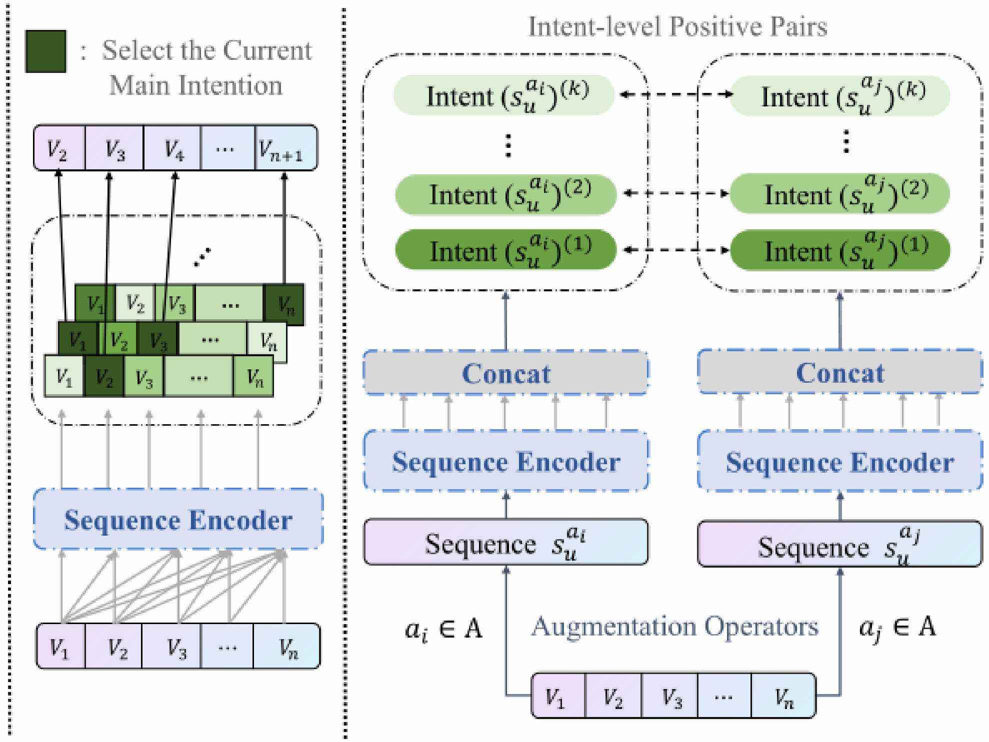
(c) Next-item Prediction (d) Contrastive Learning

Multi-Intention Contrastive Learning

$$view1(s_u^{a_i})^{(k)} = LI(s_u^{a_i})^{(k)} + GI(s_u^{a_i})^{(k)},$$
$$view2(s_u^{a_j})^{(k)} = LI(s_u^{a_j})^{(k)} + GI(s_u^{a_j})^{(k)}, \quad (13)$$

$$\mathcal{L}_{\mathrm{CL}}\left((s_u^{a_i})^{(k)}, (s_u^{a_j})^{(k)}\right) = -\log \frac{\exp\left(\mathrm{sim}\left((s_u^{a_i})^{(k)}, (s_u^{a_j})^{(k)}\right)\right)}{\sum_{s \in neg} \exp\left(\mathrm{sim}\left((s_u^{a_i})^{(k)}, s\right)\right)}. \quad (14)$$

Method



(c) Next-item Prediction (d) Contrastive Learning

Multi-Task Training

$$\mathcal{L}_{\mathrm{SR}}\left(s_u^t\right) = -\log \frac{\max_{k\in\{1,2,\ldots,k\}}\left(\exp\left(s_u^{t\top}\cdot v_{t+1}^{+}\right)\right)}{\sum_{v'_{t+1}\in\mathcal{V}} \max_{k\in\{1,2,\ldots,k\}}\left(\exp\left(s_u^{t\top}\cdot v'_{t+1}\right)\right)}, \tag{15}$$

$$\mathcal{L}_{\text{joint}} = \mathcal{L}_{\mathrm{SR}} + \lambda \mathcal{L}_{\mathrm{CL}}. \tag{16}$$

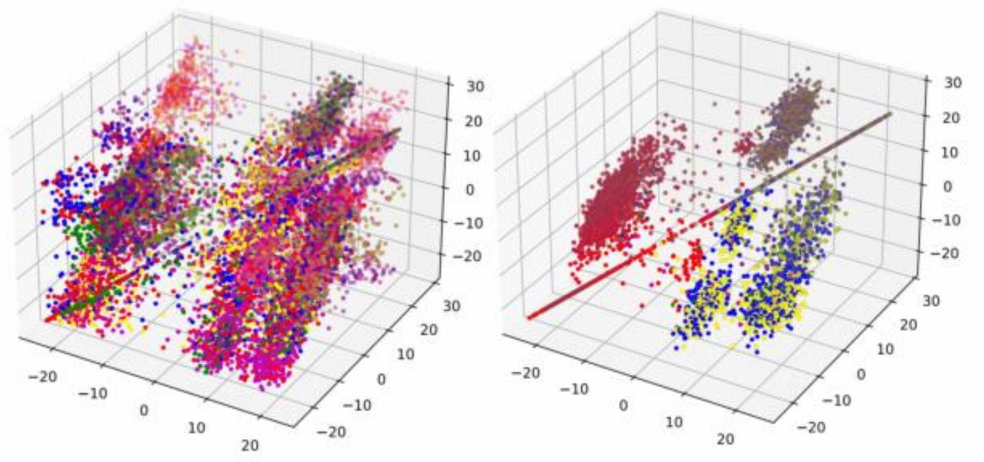
Experiments

Table 2: Performance comparison of different methods on four datasets, where our approach IOCRec's best results are in bold. The underlined numbers are the best results besides IOCRec. The reported result of IOCRec for each dataset is the best result of applying intention k.

Dataset	Metric	PopRec	GRU4Rec	Caser	BERT4Rec	SASRec	DSSRec	S^3-$Rec_{MIP,SP}$	CL4SRec	CoSeRec	$IOCRec_{CL4S}$	$IOCRec_{CoSe}$	Improv.
Sports	NDCG@5	0.1538	0.2126	0.2020	0.2341	0.2497	<u>0.2627</u>	0.2594	0.2544	0.2543	**0.2885**	0.2856	9.82%
	NDCG@10	0.1902	0.2527	0.2390	0.2775	0.2869	0.2997	<u>0.3035</u>	0.2916	0.2927	**0.3272**	0.3249	7.81%
	HR@5	0.2293	0.3055	0.2866	0.3375	0.3466	0.3617	<u>0.3673</u>	0.3518	0.3510	**0.3950**	0.3915	7.54%
	HR@10	0.3423	0.4299	0.4014	0.4722	0.4622	0.4802	<u>0.4933</u>	0.4674	0.4699	**0.5152**	0.5130	4.44%
Beauty	NDCG@5	0.1391	0.2268	0.2219	0.2622	0.2848	<u>0.2992</u>	0.2657	0.2888	0.2887	**0.3215**	0.3202	7.45%
	NDCG@10	0.1803	0.2584	0.2512	0.2975	0.3156	<u>0.3220</u>	0.3018	0.3194	0.3202	**0.3535**	0.3511	9.78%
	HR@5	0.2105	0.3125	0.3032	0.3640	0.3741	<u>0.3874</u>	0.3682	0.3779	0.3774	**0.4166**	0.4153	7.54%
	HR@10	0.3386	0.4106	0.3942	0.4739	0.4696	0.4756	<u>0.4805</u>	0.4732	0.4751	**0.5161**	0.5112	7.41%
Yelp	NDCG@5	0.1622	0.3784	0.3696	<u>0.4252</u>	0.4113	0.4231	0.3634	0.4130	0.4183	**0.4662**	0.4659	9.64%
	NDCG@10	0.2007	0.4375	0.4198	<u>0.4778</u>	0.4642	0.4711	0.4268	0.4669	0.4718	0.5162	**0.5168**	8.16%
	HR@5	0.2415	0.5437	0.5111	<u>0.5976</u>	0.5745	0.5827	0.5256	0.5772	0.5836	0.6336	**0.6365**	6.51%
	HR@10	0.3609	0.7265	0.6661	<u>0.7597</u>	0.7373	0.7412	0.7233	0.7433	0.7483	0.7872	**0.7875**	3.66%
Toys	NDCG@5	0.1286	0.1919	0.1885	0.2327	0.2820	<u>0.2934</u>	0.2307	0.2859	0.2854	**0.3152**	0.3144	7.43%
	NDCG@10	0.1618	0.2274	0.2183	0.2698	0.3136	<u>0.3256</u>	0.2742	0.3173	0.3166	**0.3464**	0.3455	6.39%
	HR@5	0.1977	0.2795	0.2614	0.3344	0.3682	0.3723	0.3368	<u>0.3749</u>	0.3735	0.4071	**0.4078**	8.78%
	HR@10	0.3008	0.3896	0.3540	0.4493	0.4663	<u>0.4798</u>	0.4729	0.4723	0.4705	0.5032	**0.5041**	5.06%

Table 1: The statistics of the datasets.

Dataset	# Users	# Items	# Actions	Avg.length	Sparsity
Sports	25,598	18,357	296,337	8.3	99.95%
Beauty	22,363	12,101	198,502	8.9	99.73%
Yelp	30,431	20,033	316,354	10.3	99.95%
Toys	19,412	11,924	167,597	8.6	99.93%



(a) Multi-Intention without CL (b) Multi-Intention Oriented CL

Figure 3: Users' multi-intention representations comparing (a) and (b) on Yelp.

Table 3: Ablation study of IOCRec (NDCG@10).

Model	Sports	Beauty	Yelp	Toys
$\text{IOCRec}_{\text{CL4S}}$	**0.3272**	**0.3535**	**0.5162**	**0.3464**
w/o GI	0.3196	0.3394	0.5150	0.3371
w/o LI	0.3135	0.3265	0.4856	0.3239
w/o IC	0.3130	0.3233	0.5043	0.3219
$\text{IOCRec}_{\text{CoSe}}$	**0.3249**	**0.3511**	**0.5168**	**0.3455**
w/o GI	0.3177	0.3433	0.5126	0.3363
w/o LI	0.3129	0.3251	0.4818	0.3221
w/o IC	0.3105	0.3291	0.4924	0.3235
CL4SRec	0.2916	0.3194	0.4669	0.3173
CL4SRec+IC	**0.3143**	**0.3311**	**0.4981**	**0.3318**
CoSeRec	0.2927	0.3202	0.4718	0.3166
CoSeRec+IC	**0.3168**	**0.3354**	**0.5017**	**0.3307**

Experiments

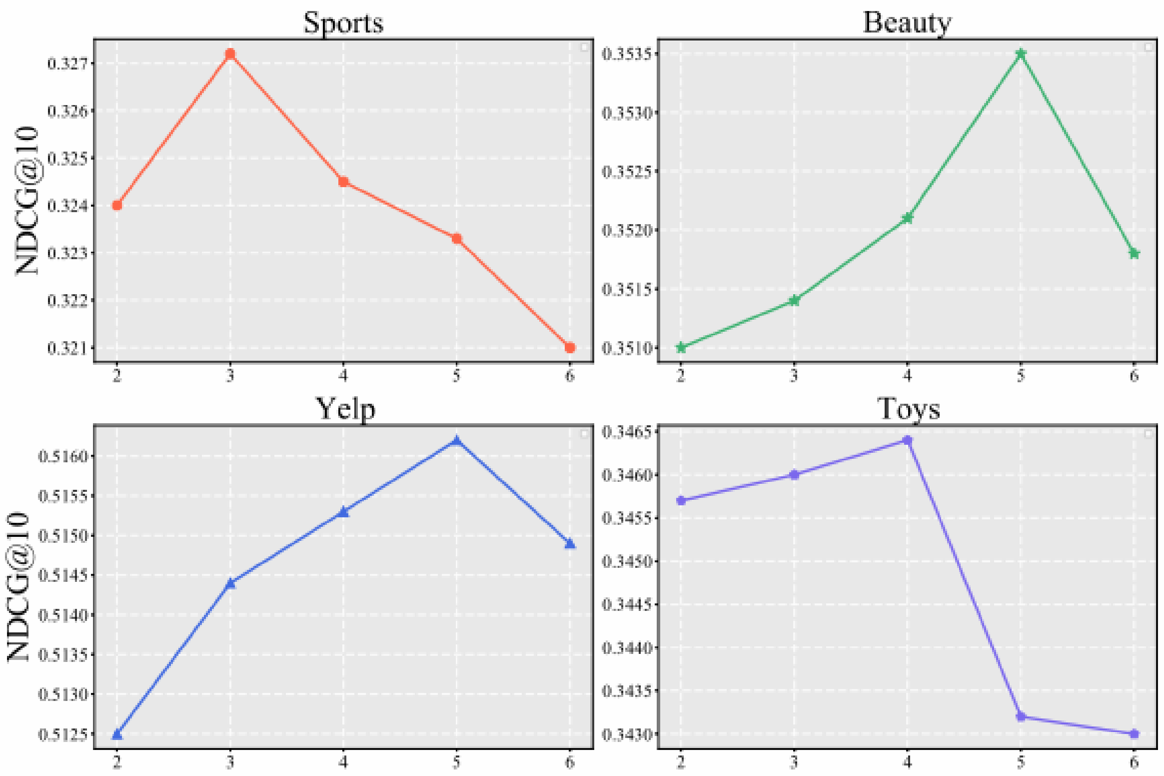


Figure 4: Performance (NDCG@10) comparison of the number k of intentions on four datasets.

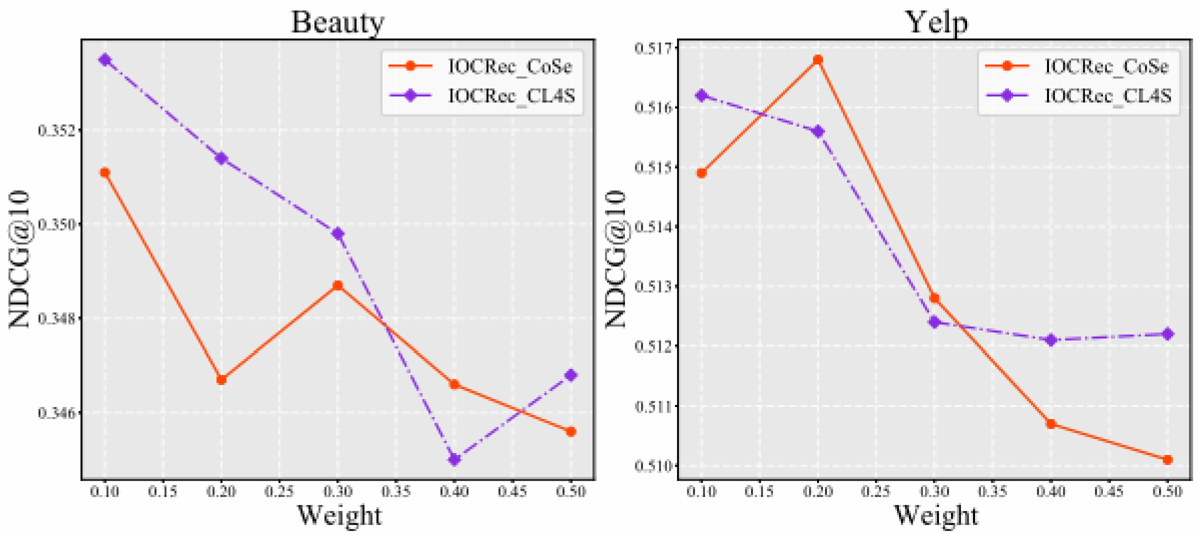


Figure 5: Performance (NDCG@10) comparison of the different λ.

Experiments

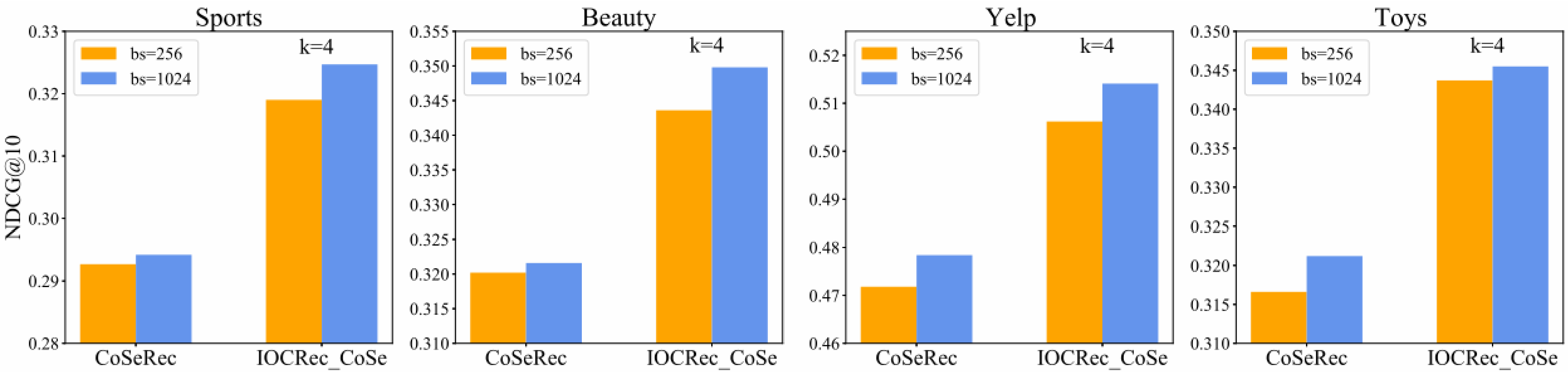


Figure 6: Performance (NDCG@10) comparison of different batch size on four datasets.