Contrastive Cross-Domain Sequential Recommendation

Jiangxia Cao
Institute of Information Engineering, Chinese Academy of Sciences
School of Cyber Security, University of Chinese Academy of Sciences
caojiangxia@iie.ac.cn

Xin Cong
Institute of Information Engineering, Chinese Academy of Sciences
School of Cyber Security, University of Chinese Academy of Sciences
congxin@iie.ac.cn

Jiawei Sheng
Institute of Information Engineering, Chinese Academy of Sciences
School of Cyber Security, University of Chinese Academy of Sciences
shengjiawei@iie.ac.cn

Tingwen Liu*
Institute of Information Engineering, Chinese Academy of Sciences
School of Cyber Security, University of Chinese Academy of Sciences
liutingwen@iie.ac.cn

Bin Wang
Xiaomi AI Lab, Xiaomi Inc
wangbin11@xiaomi.com

https://github.com/cjx96/C2DSR

——CIKM 2022

Reported by Yabo Yin

Introduction

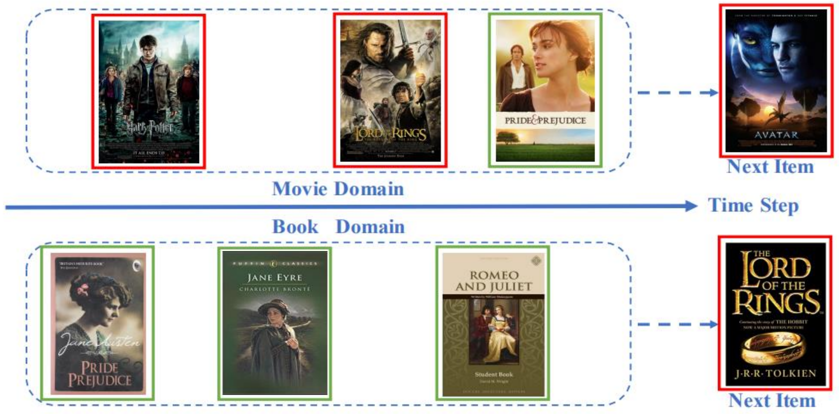


Figure 1: Illustration of user's sequential interactions in Movie and Book domains. Movie or book surrounded by the same color reflects similar user preference, where the green/red represent the "Romance"/"Fantasy" preferences.

Figure 2: A toy illustration of item relationships. The green boxes reflect the sequential pattern signal of intra-sequence item relationships. The red boxes reflect the collaborative signal of inter-sequence item relationships.

1. Simply transferring the biased single-domain preference can be intractable to describe precise cross-domain user preference, which would easily lead to unstable and sub-optimal recommendation results.

2. previous CDSR works ignore the inter-sequence relationship of items, which provides valuable collaborative signal to generate better user representation.

Method

Figure 2: A toy illustration of item relationships. The green boxes reflect the sequential pattern signal of intra-sequence item relationships. The red boxes reflect the collaborative signal of inter-sequence item relationships.

Method

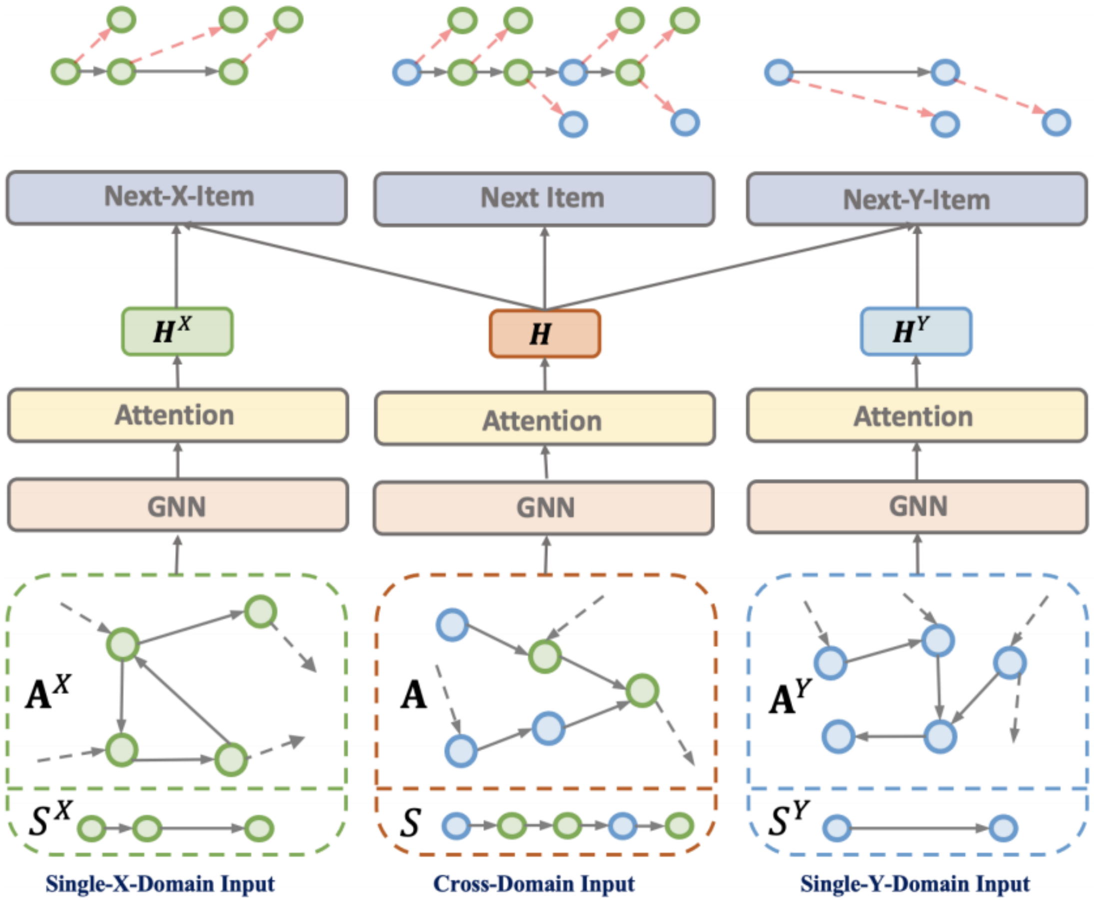


Figure 3: A toy example of sequential training objective for CDSR. The red dotted lines indicate the next prediction item.

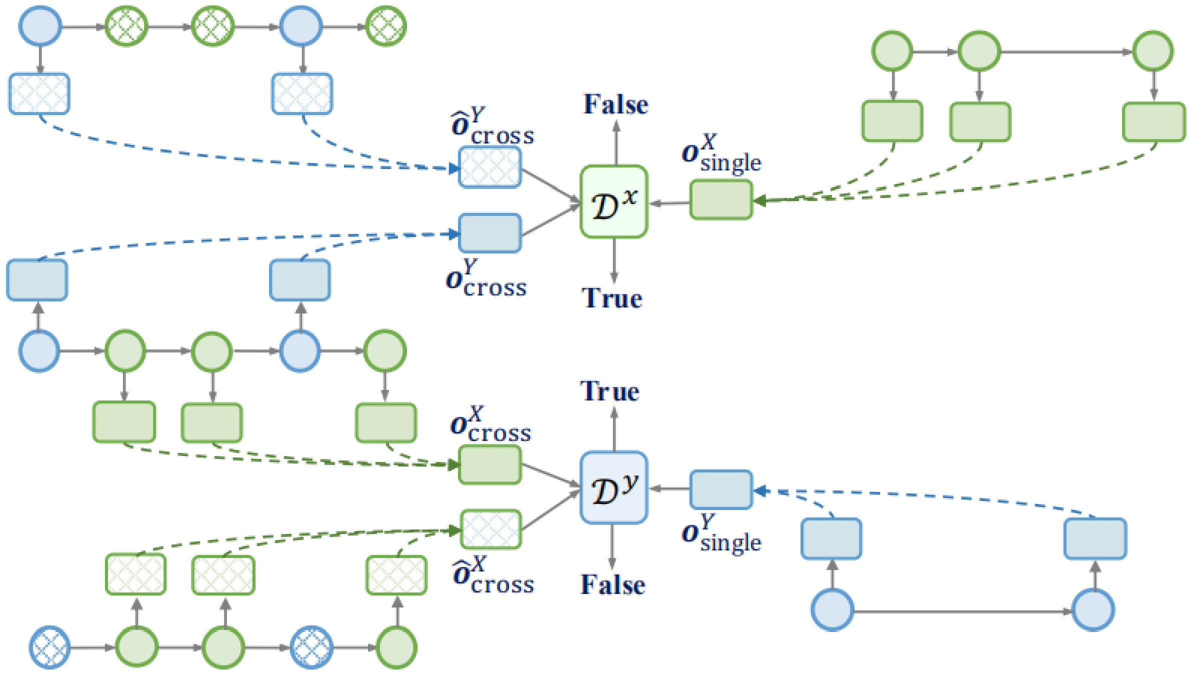


Figure 4: Illustration of our contrastive infomax. The hollow green/blue circles denote random items in domain X/Y.

Method

Preliminaries

interaction sequence involves two domains, namely domain X and domain Y.

$(S^X, S^Y, S)_u \in \mathcal{S}$ belongs to a certain user u.

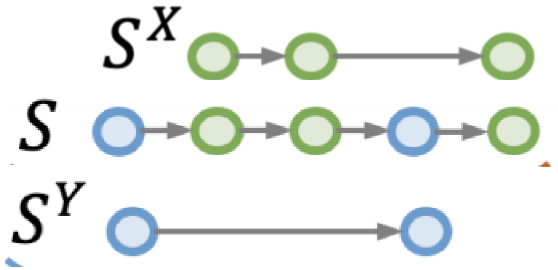
$S^X = [x_1, x_2, \ldots, x_{|S^X|}]$ and $S^Y = [y_1, y_2, \ldots, y_{|S^Y|}]$

$S = [y_1, x_1, x_2, \ldots, y_{|S^Y|}, \ldots, x_{|S^X|}]$

$\mathbf{A}^X \in \{0,1\}^{|\mathcal{X}|\times|\mathcal{X}|}, \mathbf{A}^Y \in \{0,1\}^{|\mathcal{Y}|\times|\mathcal{Y}|}, \mathbf{A} \in \{0,1\}^{(|\mathcal{X}|+|\mathcal{Y}|)\times(|\mathcal{X}|+|\mathcal{Y}|)}$

where $\mathbf{A}^X_{ij} = 1$ if x_j is the one next item of x_i,

$$\begin{aligned} &\operatorname{argmax}_{x_i \in \mathcal{X}} \mathrm{P}^X\left(x_i | S^X, S^Y, S\right), \text{if next item} \in \mathcal{X} \\ &\operatorname{argmax}_{y_j \in \mathcal{Y}} \mathrm{P}^Y\left(y_j | S^X, S^Y, S\right), \text{if next item} \in \mathcal{Y} \end{aligned} \quad (1)$$



Method

$$\mathbf{E}^X \in \mathbb{R}^{|\mathcal{X}|\times d}, \quad \mathbf{E}^Y \in \mathbb{R}^{|\mathcal{Y}|\times d} \quad \mathbf{E} \in \mathbb{R}^{(|\mathcal{X}|+|\mathcal{Y}|)\times d}, \quad \mathbf{T} \in \mathrm{R}^{M\times d}$$

$$\mathbf{G}_0^X = \mathbf{E}^X, \mathbf{G}_0^Y = \mathbf{E}^Y, \mathbf{G}_0 = \mathbf{E},$$

$$\mathbf{G}_1^X = \mathtt{Norm}(\mathbf{A}^X)\mathbf{G}_0^X, \quad \mathbf{G}_1^Y = \mathtt{Norm}(\mathbf{A}^Y)\mathbf{G}_0^Y, \quad \mathbf{G}_1 = \mathtt{Norm}(\mathbf{A})\mathbf{G}_0, \qquad (2)$$

$\mathtt{Norm}(\cdot)$ denote the row-normalized function

$$\mathbf{G}^X = \mathtt{Mean}(\mathbf{G}_l^X) + \mathbf{E}^X, \mathbf{G}^Y = \mathtt{Mean}(\mathbf{G}_l^Y) + \mathbf{E}^Y, \mathbf{G} = \mathtt{Mean}(\mathbf{G}_l) + \mathbf{E}. \qquad (3)$$

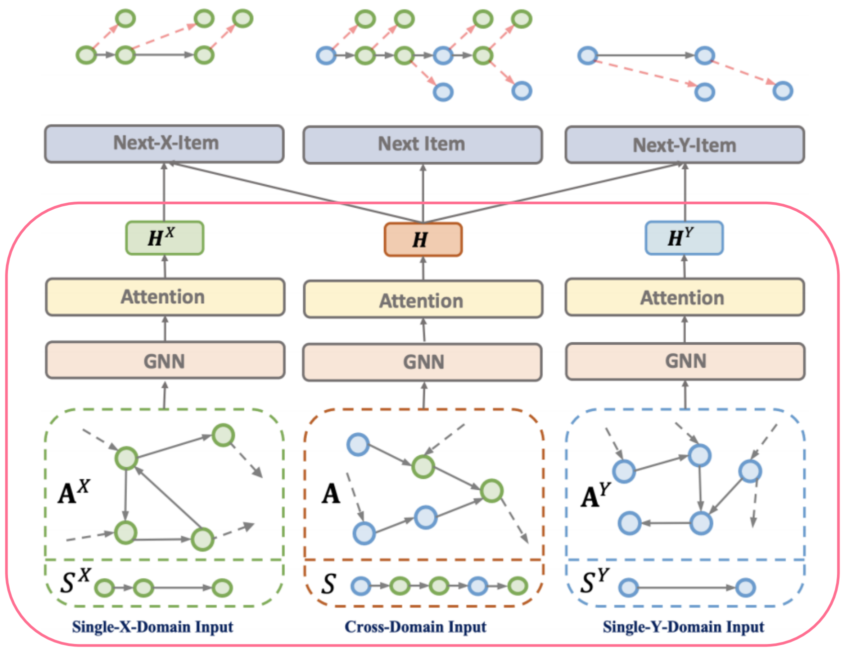


Figure 3: A toy example of sequential training objective for CDSR. The red dotted lines indicate the next prediction item.

$$S^X = [\text{<pad>}, x_1, x_2, \text{<pad>}, x_3] \quad S^Y = [y_1, \text{<pad>}, \text{<pad>}, y_2, \text{<pad>}] \quad S = [y_1, x_1, x_2, y_2, x_3]$$

$$\begin{aligned} \boldsymbol{H}^X &= \mathtt{AttEncoder}^X(S^X, \mathbf{G}^X), \quad \boldsymbol{H}^Y = \mathtt{AttEncoder}^Y(S^Y, \mathbf{G}^Y), \\ \boldsymbol{H} &= \mathtt{AttEncoder}(S, \mathbf{G}), \end{aligned} \qquad (4)$$

$$\boldsymbol{H}^X \in \mathbb{R}^{|S|\times d}, \boldsymbol{H}^Y \in \mathbb{R}^{|S|\times d}, \boldsymbol{H} \in \mathbb{R}^{|S|\times d}$$

Method

Single-Domain Item Prediction.

$$\mathcal{L}^X_{\text{single}} = \sum_{S^X \in \mathcal{S}} \sum_t \mathcal{L}^X_{\text{single}}(S^X, t)$$

$$\mathcal{L}^X_{\text{single}}(S^X, t) = -\log \mathrm{P}^X_{\text{single}}(x_{t+1} | [\text{<pad>}, x_1, x_2, \text{<pad>}, \ldots, x_t]), \tag{5}$$

$$\mathrm{P}^X_{\text{single}}(x_{t+1} | [\ldots, x_t]) = \mathtt{Softmax}(\boldsymbol{h}^X_t \mathbf{W}^X + \boldsymbol{h}_t \mathbf{W}^X)_{x_{t+1}} \tag{6}$$

$$\boldsymbol{h}^X_t \in \mathbb{R}^{1\times d}, \boldsymbol{h}_t \in \mathbb{R}^{1\times d} \qquad \mathbf{W}^X \in \mathbb{R}^{d\times|\mathcal{X}|}$$

Cross-Domain Item Prediction.

$$\mathcal{L}_{\text{cross}} = \sum_{S \in \mathcal{S}} \sum_t \mathcal{L}_{\text{cross}}(S, t),$$

$$\mathcal{L}_{\text{cross}}(S, t) = \begin{cases} -\log \mathrm{P}^X_{\text{cross}}(x_{t+1} | [y_1, x_1, x_2, \ldots, x_t]), \\ -\log \mathrm{P}^Y_{\text{cross}}(y_{t+1} | [y_1, x_1, x_2, \ldots, x_t]), \end{cases} \tag{7}$$

$$\begin{aligned} \mathrm{P}^X_{\text{cross}}(x_{t+1} | [y_1, x_1, x_2, \ldots, x_t]) &= \mathtt{Softmax}(\boldsymbol{h}_t \mathbf{W}^X)_{x_{t+1}}, \\ \mathrm{P}^Y_{\text{cross}}(y_{t+1} | [y_1, x_1, x_2, \ldots, x_t]) &= \mathtt{Softmax}(\boldsymbol{h}_t \mathbf{W}^Y)_{y_{t+1}}, \end{aligned} \tag{8}$$

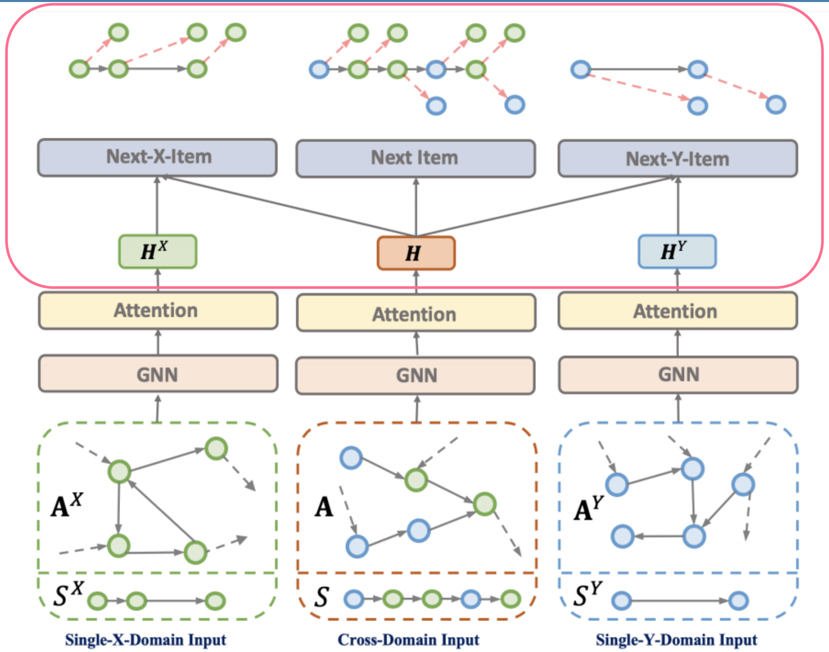


Figure 3: A toy example of sequential training objective for CDSR. The red dotted lines indicate the next prediction item.

Method

Single- and Cross- Domain Prototype Representations.

$$\boldsymbol{o}^X_{\text{single}} = \mathsf{Mean}(\boldsymbol{H}^X), \qquad \boldsymbol{o}^Y_{\text{single}} = \mathsf{Mean}(\boldsymbol{H}^Y), \tag{9}$$

$$\boldsymbol{o}^X_{\text{single}} \in \mathbb{R}^{1\times d} \text{ and } \boldsymbol{o}^Y_{\text{single}} \in \mathbb{R}^{1\times d}$$

$$\boldsymbol{o}^X_{\text{cross}} = \mathsf{Mean}(\{\boldsymbol{h}_t : S_t \in \mathcal{X}\}), \quad \boldsymbol{o}^Y_{\text{cross}} = \mathsf{Mean}(\{\boldsymbol{h}_t : S_t \in \mathcal{Y}\}), \tag{10}$$

$$\begin{aligned} \widehat{S}^X &= \mathsf{Corrupt}^X(S) = [\widehat{y}_1, x_1, x_2, \widehat{y}_2, \dots], \\ \widehat{S}^Y &= \mathsf{Corrupt}^Y(S) = [y_1, \widehat{x}_1, \widehat{x}_2, y_2, \dots], \end{aligned} \tag{11}$$

$$\mathcal{L}^X_{\text{disc}} = \sum_{(S^X, S^Y, S)_u \in \mathcal{S}} -\Big(\log \mathcal{D}^X(\boldsymbol{o}^X_{\text{single}}, \boldsymbol{o}^Y_{\text{cross}}) + \log\big(1 - \mathcal{D}^X(\boldsymbol{o}^X_{\text{single}}, \widehat{\boldsymbol{o}}^Y_{\text{cross}})\big)\Big) \tag{12}$$

$$\mathcal{D}^X(\boldsymbol{o}^X_{\text{single}}, \boldsymbol{o}^Y_{\text{cross}}) = \sigma\big(\boldsymbol{o}^X_{\text{single}} \mathbf{W}^X_{\text{disc}} (\boldsymbol{o}^Y_{\text{cross}})^\top\big), \tag{13}$$

$$\mathbf{W}^X_{\text{disc}} \in \mathbb{R}^{d\times d}$$

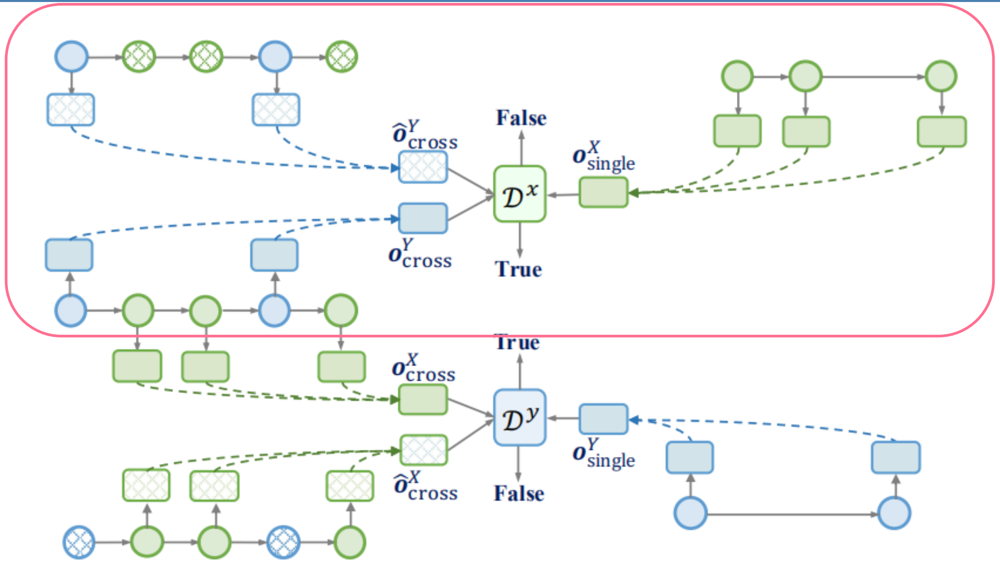


Figure 4: Illustration of our contrastive infomax. The hollow green/blue circles denote random items in domain X/Y.

$$\mathcal{L} = \underbrace{\lambda(\mathcal{L}_{\text{cross}} + \mathcal{L}^X_{\text{single}} + \mathcal{L}^Y_{\text{single}})}_{\text{Sequential training objective}} + \underbrace{(1-\lambda)(\mathcal{L}^X_{\text{disc}} + \mathcal{L}^Y_{\text{disc}})}_{\text{Contrastive infomax objective}} \tag{14}$$

$$\begin{aligned} &\operatorname{argmax}_{x_i \in \mathcal{X}} \mathrm{P}^X\left(x_i | S^X, S^Y, S\right), \quad \text{where} \\ &\mathrm{P}^X(x_i | S^X, S^Y, S) = \mathsf{Softmax}(\boldsymbol{h}^X_{|S|}\mathbf{W}^X + \boldsymbol{h}_{|S|}\mathbf{W}^X)_{x_i}. \end{aligned} \tag{15}$$

Method

$$\mathcal{L} = \underbrace{\lambda(\mathcal{L}_{\text{cross}} + \mathcal{L}^X_{\text{single}} + \mathcal{L}^Y_{\text{single}})}_{\text{Sequential training objective}} + \underbrace{(1-\lambda)(\mathcal{L}^X_{\text{disc}} + \mathcal{L}^Y_{\text{disc}})}_{\text{Contrastive infomax objective}} \quad (14)$$

In the evaluation stage,

$$\begin{aligned} &\operatorname{argmax}_{x_i \in \mathcal{X}} \mathrm{P}^X\left(x_i | S^X, S^Y, S\right), \quad \text{where} \\ &\mathrm{P}^X(x_i | S^X, S^Y, S) = \mathtt{Softmax}(\boldsymbol{h}^X_{|S|}\mathbf{W}^X + \boldsymbol{h}_{|S|}\mathbf{W}^X)_{x_i}. \end{aligned} \quad (15)$$

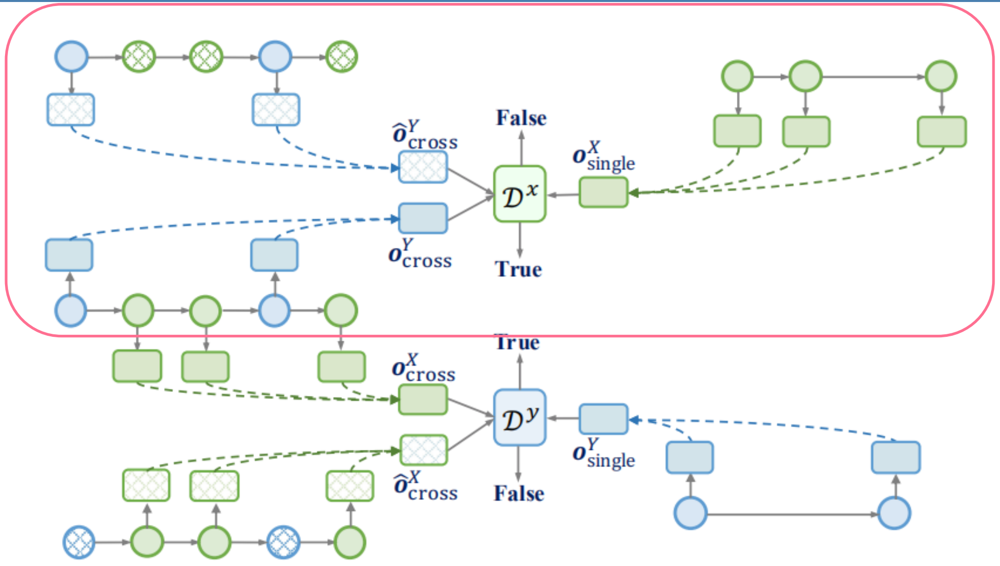


Figure 4: Illustration of our contrastive infomax. The hollow green/blue circles denote random items in domain X/Y.

Method

$$\mathbf{o}^X_{\text{single}} = \text{Mean}(\boldsymbol{H}^X), \qquad \mathbf{o}^Y_{\text{single}} = \text{Mean}(\boldsymbol{H}^Y), \tag{9}$$

$$\mathbf{o}^X_{\text{cross}} = \text{Mean}(\{\boldsymbol{h}_t : S_t \in \mathcal{X}\}), \quad \mathbf{o}^Y_{\text{cross}} = \text{Mean}(\{\boldsymbol{h}_t : S_t \in \mathcal{Y}\}), \tag{10}$$

$$\begin{aligned} \widehat{S}^X &= \text{Corrupt}^X(S) = [\widehat{y}_1, x_1, x_2, \widehat{y}_2, \ldots], \\ \widehat{S}^Y &= \text{Corrupt}^Y(S) = [y_1, \widehat{x}_1, \widehat{x}_2, y_2, \ldots], \end{aligned} \tag{11}$$

$$\mathcal{L}^X_{\text{disc}} = \sum_{(S^X, S^Y, S)_u \in \mathcal{S}} -\left(\log \mathcal{D}^X(\mathbf{o}^X_{\text{single}}, \mathbf{o}^Y_{\text{cross}}) + \log\left(1 - \mathcal{D}^X(\mathbf{o}^X_{\text{single}}, \widehat{\mathbf{o}}^Y_{\text{cross}})\right)\right) \tag{12}$$

$$\mathcal{D}^X(\mathbf{o}^X_{\text{single}}, \mathbf{o}^Y_{\text{cross}}) = \sigma(\mathbf{o}^X_{\text{single}} \mathbf{W}^X_{\text{disc}} (\mathbf{o}^Y_{\text{cross}})^\top), \tag{13}$$

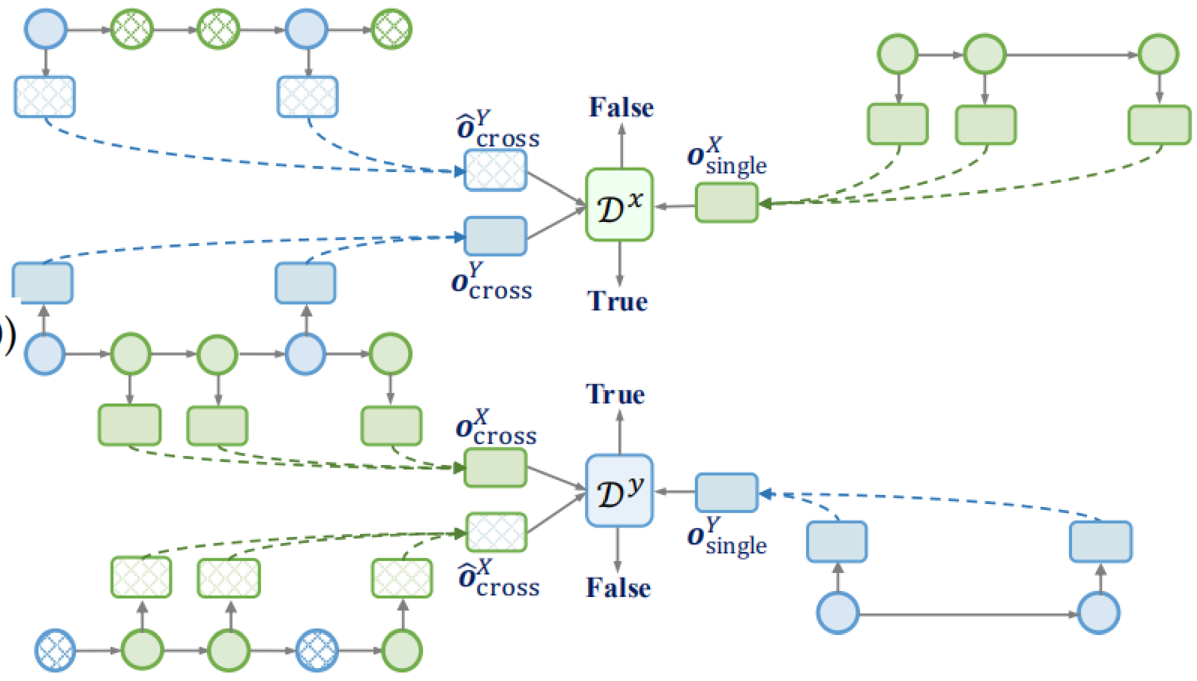


Figure 4: Illustration of our contrastive infomax. The hollow
n/blue circles denote random items in domain X/Y.

Method

$$\mathcal{L} = \underbrace{\lambda(\mathcal{L}_{\text{cross}} + \mathcal{L}^X_{\text{single}} + \mathcal{L}^Y_{\text{single}})}_{\text{Sequential training objective}} + \underbrace{(1-\lambda)(\mathcal{L}^X_{\text{disc}} + \mathcal{L}^Y_{\text{disc}})}_{\text{Contrastive infomax objective}} \tag{14}$$

$$\operatorname{argmax}_{x_i \in \mathcal{X}} \mathrm{P}^X\left(x_i | S^X, S^Y, S\right), \quad \text{where}$$
$$\mathrm{P}^X(x_i | S^X, S^Y, S) = \mathtt{Softmax}\left(\boldsymbol{h}^X_{|S|}\mathbf{W}^X + \boldsymbol{h}_{|S|}\mathbf{W}^X\right)_{x_i}. \tag{15}$$

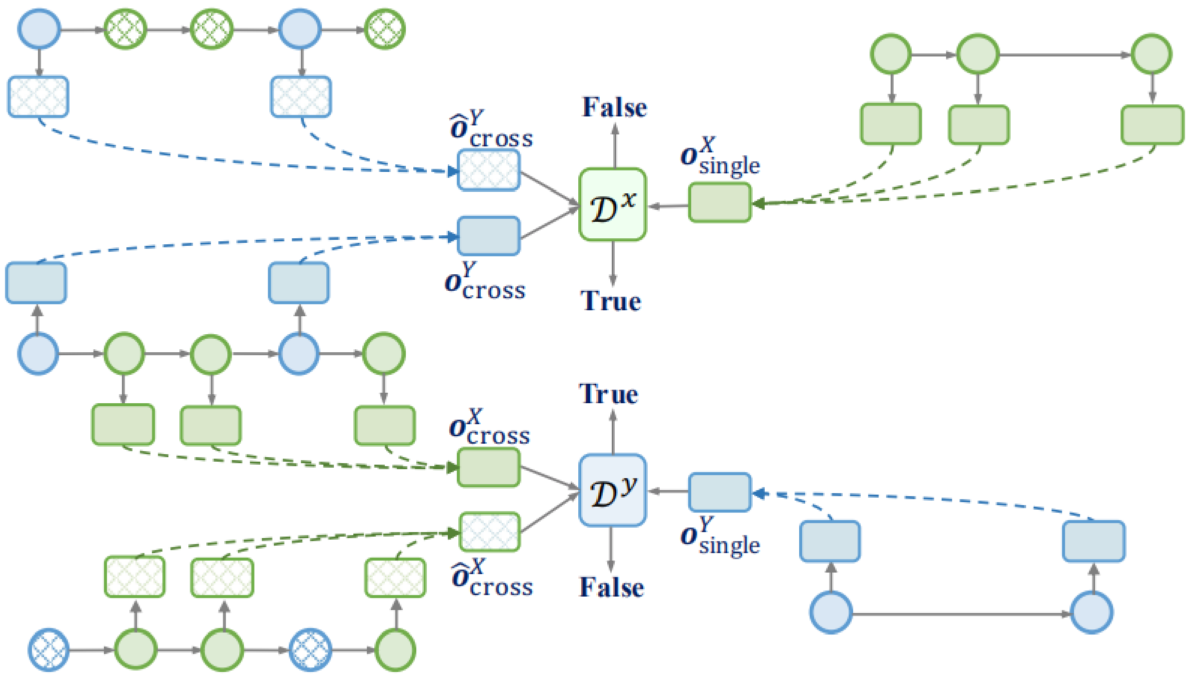


Figure 4: Illustration of our contrastive infomax. The hollow green/blue circles denote random items in domain X/Y.

Experiments

Table 1: Statistics of Three CDSR scenarios.

Scenarios	#Items	#Train	#Valid	#Test	Avg.length
Food Kitchen	29,207 34,886	34,117	2,722 5,451	2,747 5,659	9.91
Movie Book	36,845 63,937	58,515	2,032 5,612	1,978 5,730	11.98
Entertainment Education	8,367 11,404	120,635	4,525 2,404	4,485 2,300	29.94

Experiments

Table 2: Experimental results (%) on the Food-Kitchen scenario.

Methods	Food-domain recommendation						Kitchen-domain recommendation					
	MRR	NDCG		HR			MRR	NDCG		HR		
		@5	@10	@1	@5	@10		@5	@10	@1	@5	@10
BPRMF	4.10	3.55	4.03	2.42	4.51	5.95	2.01	1.45	1.85	0.73	2.18	3.43
ItemKNN	3.92	3.51	3.97	2.41	4.59	5.98	1.89	1.28	1.75	0.58	1.99	3.26
NCF-MLP	4.49	3.94	4.51	2.68	5.10	6.86	2.18	1.57	2.03	0.91	2.23	3.65
CoNet	4.13	3.61	4.14	2.42	4.77	6.35	2.17	1.50	2.11	0.95	2.07	3.71
GRU4Rec	5.79	5.48	6.13	3.63	7.12	9.11	3.06	2.55	3.10	1.61	3.50	5.22
SASRec	7.30	6.90	7.79	4.73	8.92	11.68	3.79	3.35	3.93	1.92	4.78	6.62
SR-GNN	7.84	7.58	8.35	5.03	9.88	12.27	4.01	3.47	4.13	2.07	4.80	6.84
π-Net	7.68	7.32	8.13	5.25	9.25	11.75	3.53	2.98	3.73	1.57	4.34	6.67
PSJNet	8.33	8.07	8.77	5.73	10.28	12.45	4.10	3.68	4.32	2.14	5.17	7.15
MIFN	8.55	8.28	9.01	6.02	10.43	12.71	4.09	3.57	4.29	2.21	4.86	7.08
C^2DSR	**8.91**	**8.65**	**9.71**	**5.84**	**11.24**	**14.54**	**4.65**	**4.16**	**4.94**	**2.51**	**5.74**	**8.18**

Experiments

Table 3: Experimental results (%) on the Movie-Book scenario.

Methods	Movie-domain recommendation						Book-domain recommendation					
	MRR	NDCG		HR			MRR	NDCG		HR		
		@5	@10	@1	@5	@10		@5	@10	@1	@5	@10
BPRMF	2.96	2.18	2.80	1.41	3.03	4.95	1.27	0.85	1.17	0.48	1.23	2.25
ItemKNN	2.92	2.17	2.88	1.26	3.13	5.35	1.26	0.86	1.10	0.48	1.25	2.00
NCF-MLP	3.05	2.26	2.96	1.41	3.13	5.30	1.43	1.06	1.26	0.62	1.39	2.18
CoNet	3.07	2.42	3.01	1.31	3.48	5.35	1.45	1.04	1.28	0.64	1.44	2.19
GRU4Rec	3.83	3.14	3.73	2.27	3.39	5.40	1.68	1.34	1.52	0.91	1.81	2.37
SASRec	3.79	3.23	3.69	2.37	3.99	5.20	1.81	1.41	1.71	0.95	1.83	2.75
SR-GNN	3.85	3.27	3.78	2.22	4.19	5.81	1.78	1.40	1.66	0.89	1.90	2.72
π-Net	4.16	3.72	4.17	2.52	4.75	6.11	2.17	1.84	2.03	1.43	2.25	2.84
PSJNet	4.63	4.06	4.76	2.78	5.30	7.53	2.44	2.07	2.35	1.66	2.58	3.28
MIFN	5.05	4.21	5.20	2.83	5.51	8.29	2.51	2.12	2.31	1.60	2.46	3.07
C^2DSR	**5.54**	**4.76**	**5.76**	**3.13**	**6.47**	**9.55**	**2.55**	**2.17**	**2.45**	**1.71**	**2.84**	**3.75**

Experiments

Table 4: Experimental results (%) on the Entertainment-Education scenario.

Methods	Entertainment-domain recommendation						Education-domain recommendation					
	MRR	NDCG		HR			MRR	NDCG		HR		
		@5	@10	@1	@5	@10		@5	@10	@1	@5	@10
BPRMF	45.97	47.38	50.11	35.98	57.65	66.08	46.50	47.51	49.27	38.69	54.86	60.26
ItemKNN	47.81	49.24	52.01	37.91	59.44	67.93	46.22	47.23	49.22	38.08	54.86	61.04
NCF-MLP	44.94	47.57	50.61	31.66	61.73	71.08	46.24	48.73	50.66	35.17	60.56	66.43
CoNet	45.76	48.31	51.63	32.30	62.45	72.68	47.83	50.11	52.22	36.60	61.65	68.04
GRU4Rec	45.61	47.48	51.46	32.73	61.06	73.35	51.35	53.88	56.21	39.95	66.13	73.21
SASRec	50.44	52.67	56.10	37.05	66.39	76.92	53.69	55.87	58.64	41.86	68.04	76.56
SR-GNN	50.67	52.77	56.47	37.43	66.30	77.17	54.74	56.73	59.69	43.08	68.43	77.47
π-Net	52.68	54.88	57.63	40.91	67.15	75.11	55.05	57.23	59.32	44.26	68.21	74.65
PSJNet	53.50	**57.57**	**60.07**	**42.52**	68.94	76.56	55.94	58.18	60.15	**45.21**	69.01	75.08
C^2DSR	**53.87**	56.20	59.35	40.89	**69.45**	**79.08**	**56.72**	**59.05**	**61.56**	45.13	**71.04**	**79.21**

Experiments

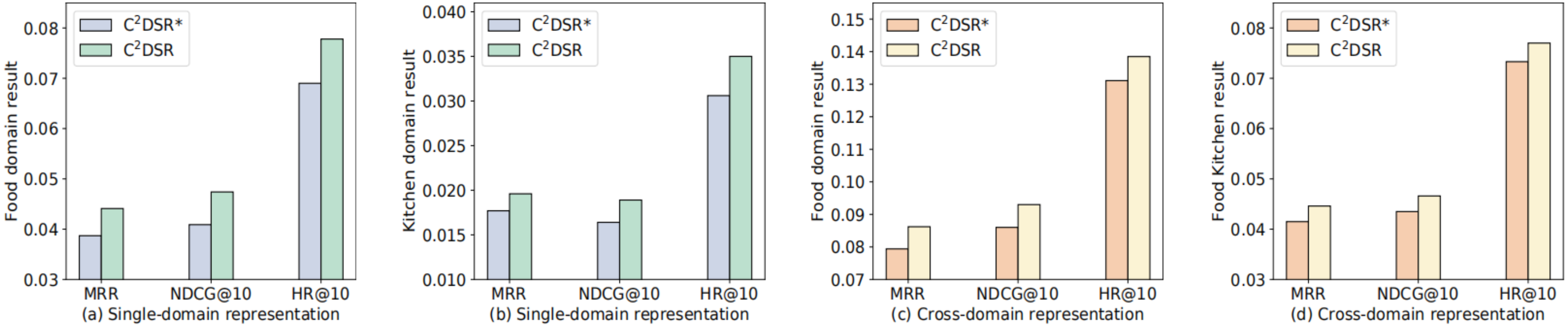


Figure 5: The predictive results of single-domain and cross-domain representations on Food-Kitchen.

Experiments

Table 5: Variants results on Food-Kitchen.

Model Variants	Food-domain			Kitchen-domain		
	MRR	NDCG@10	HR@10	MRR	NDCG@10	HR@10
GRU4Rec	5.79	6.13	9.11	3.06	3.10	5.22
GRU4Rec++	6.92	7.18	10.10	3.21	3.28	5.63
C^2DSR(GRU4Rec)*	6.83	7.26	10.72	3.57	3.68	6.10
C^2DSR(GRU4Rec)	7.68	8.02	11.60	3.92	4.02	6.54
SASRec	7.30	7.79	11.68	3.79	3.93	6.62
SASRec++	8.09	8.70	13.11	4.20	4.45	7.55
C^2DSR*	8.73	9.36	13.66	4.41	4.55	7.70
C^2DSR	8.91	9.71	14.54	4.65	4.94	8.18

Experiments

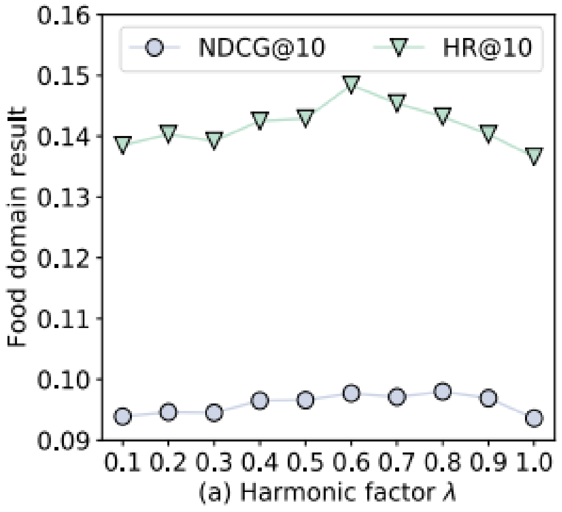


Figure 6: Result of harmonic factor λ.

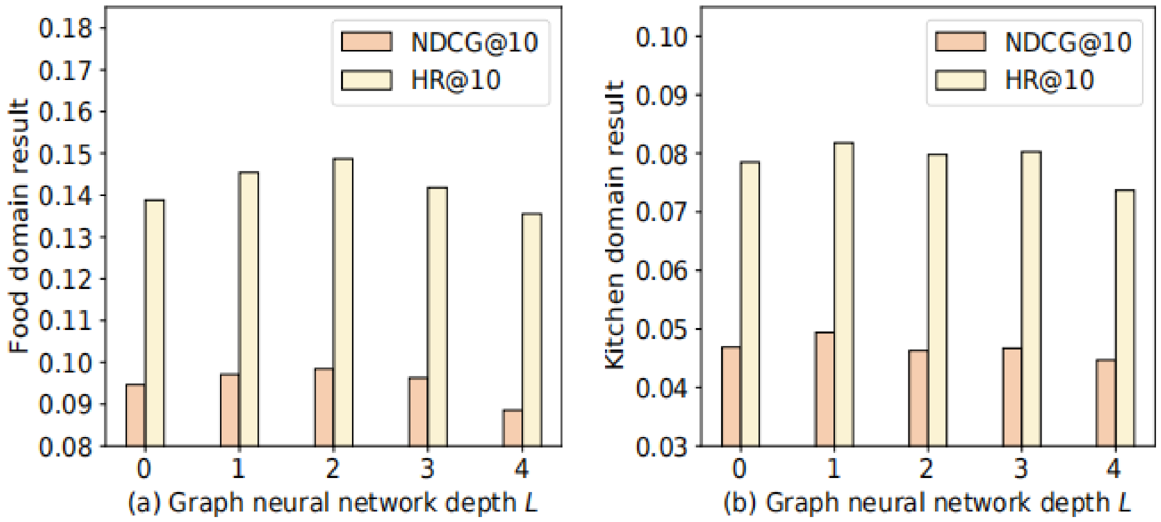


Figure 7: Impact of GNN depth L.

Thank you!