Towards Universal Cross-Domain Recommendation

Advisor :Jia-Ling, Koh

Speaker : Shu-Ming Yu

Source : WSDM’ 23

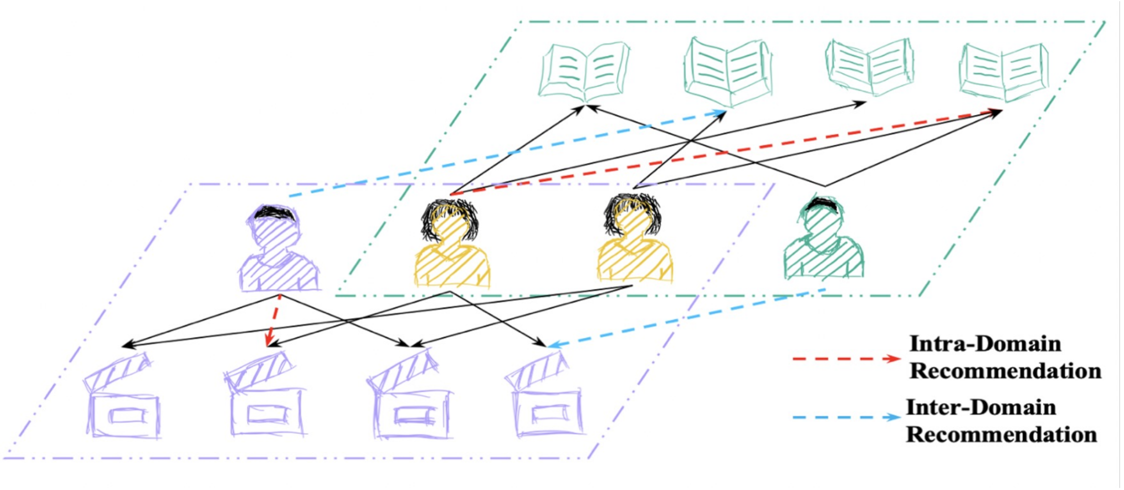
Date : 2023/08/29

Outline

- **Introduction**
- Method
- Experiment
- Conclusion

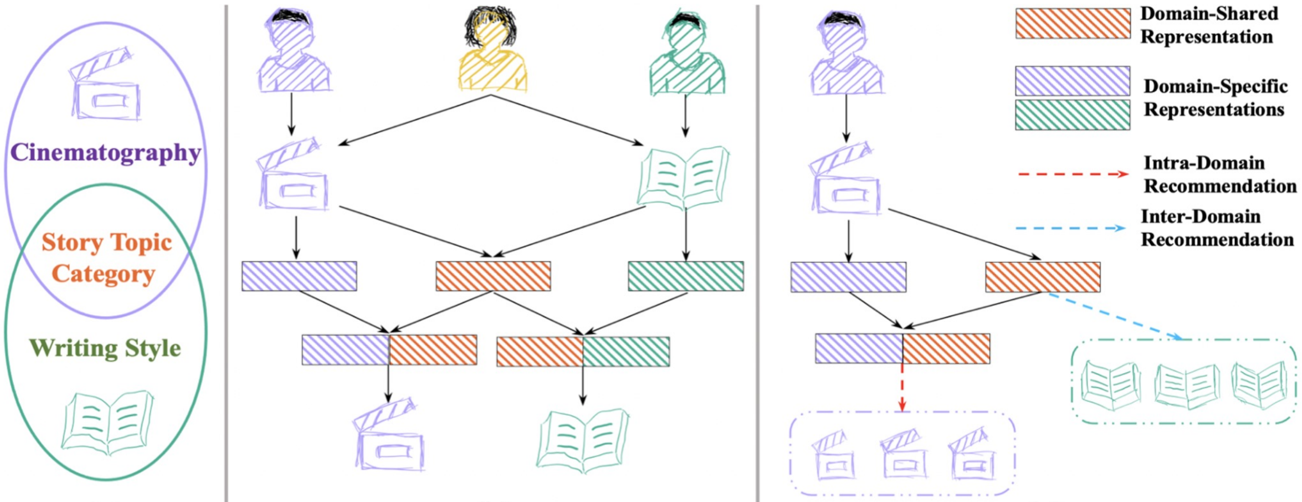
Introduction

- Cross-domain recommendation can be roughly divided into two branches.
  - Intra-domain (user with few interactions)
  - Inter-domain (new-coming user)



Introduction

- Domain-shared information can help the prediction of both the source domain and the target domain.



(a) (b) (c)

Data Notation

Input:

- Domain $\{\mathcal{D}^X, \mathcal{D}^Y\}$
  - $\mathcal{D} = (\mathcal{U}, \mathcal{V}, \mathcal{E})$

    $\mathcal{U}$: user set , $\mathcal{V}$: item set , $\mathcal{E}$: interaction set

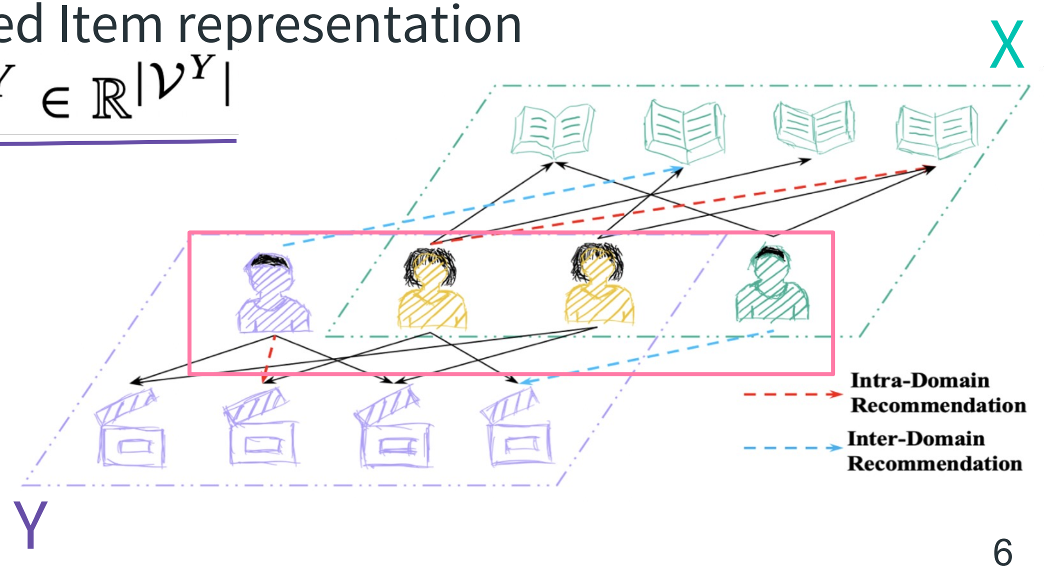
    ($\mathcal{E}$ is a binary interaction matrix $\mathbf{A} \in \{0, 1\}^{|\mathcal{U}| \times |\mathcal{V}|}$)

Output:

- interaction $y_{(u,v_i)}$, $y_{(u,v_i)} \in \{0, 1\}$

Data Notation

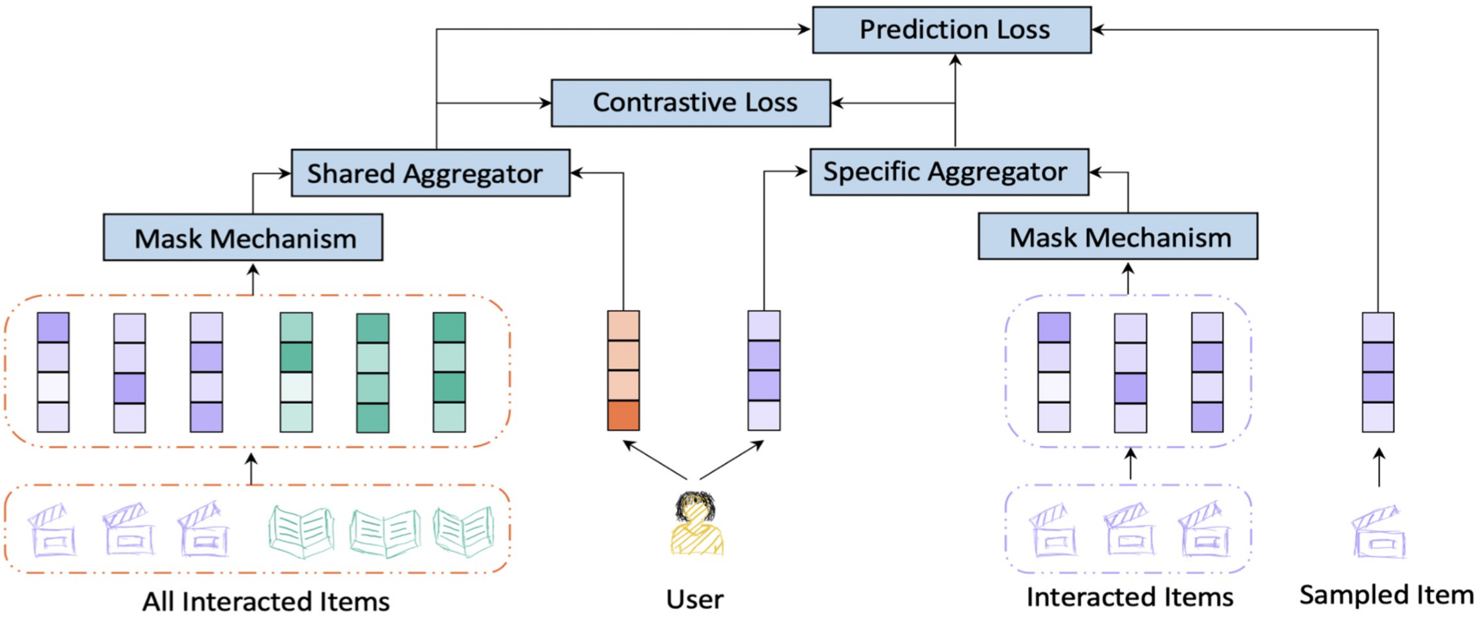
- Domain specific/shared user representation
  - $\mathbf{U}^X \in \mathbb{R}^{|\mathcal{U}^X|}$ $\mathbf{U}^Y \in \mathbb{R}^{|\mathcal{U}^Y|}$ $\mathbf{U}^S \in \mathbb{R}^{|\mathcal{U}^X \cup \mathcal{U}^Y|}$
- Domain specific/shared Item representation
  - $\mathbf{V}^X \in \mathbb{R}^{|\mathcal{V}^X|}$, $\mathbf{V}^Y \in \mathbb{R}^{|\mathcal{V}^Y|}$



Outline

- Introduction
- **Method**
  - Masking Mechanism
  - Aggregator Architecture
  - Contrastive loss & Prediction loss
- Experiment
- Conclusion

Method



Prediction Loss
Contrastive Loss
Shared Aggregator
Specific Aggregator
Mask Mechanism
Mask Mechanism
All Interacted Items
User
Interacted Items
Sampled Item

Masking Mechanism

- A user's historically interacted items

$$\mathcal{H}_u^X = \{v_1^X, v_2^X, v_3^X, \dots\} \ / \ \mathcal{H}_u^Y = \{v_1^Y, v_2^Y, v_3^Y, \dots\}$$

$$\mathcal{H}_u^S = \{\mathcal{H}_u^X, \mathcal{H}_u^Y\}$$

- Interaction masking:

$$\widetilde{\mathcal{H}}_u = \textsc{Mask}(\mathcal{H}_u, \boldsymbol{m}), \text{ where } \boldsymbol{m}_i = \textsc{Bernoulli}(p)$$

- Domain masking:

$$\widehat{\mathcal{H}}_{u,X}^S = \{\emptyset, \widetilde{\mathcal{H}}_u^Y\}, \quad \widehat{\mathcal{H}}_{u,Y}^S = \{\widetilde{\mathcal{H}}_u^X, \emptyset\}$$

Aggregator Architecture

- **Goal** : find a well user representation

$$\mathbf{h}_u^X = \lambda_A \cdot \mathbf{u}^X + (1 - \lambda_A) \cdot \textsc{Aggregator}^X(\mathbf{u}^X, \mathcal{H}_u^X),$$

$$\mathbf{h}_u^S = \lambda_A \cdot \mathbf{u}^S + (1 - \lambda_A) \cdot \textsc{Aggregator}^S(\mathbf{u}^S, \mathcal{H}_u^S),$$

- Mean-pooling

$$\mathbf{h} = \mathrm{MEAN}(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots\})\mathbf{W}_{\text{agg}}$$

element-wise mean

Aggregator Architecture

- User-attention-pooling

$$\boldsymbol{\alpha} = \textsc{Softmax}(\{\alpha_1, \alpha_2, \alpha_3, \ldots\}), \text{ where } \alpha_i = \textsc{Tanh}(\mathbf{W}_{\mathrm{att}}\mathbf{v}_i + \mathbf{b})\mathbf{u}^\top$$

sum is 1

$$\mathbf{h} = \textsc{WeightedMean}(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots\}, \boldsymbol{\alpha})\mathbf{W}_{\mathrm{agg}}$$

Aggregator Architecture

- Item-similarity-pooling

$$\boldsymbol{\alpha} = \textsc{Normlize}(\{\alpha_1, \alpha_2, \alpha_3, \ldots\}), \text{ where } \alpha_i = \boxed{(\mathbf{AB})_{u,v_i}}$$

$$\mathbf{h} = \textsc{WeightedMEAN}(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots\}, \boldsymbol{\alpha})\mathbf{W}_{\text{agg}}$$

A : interaction matrix , $\mathbf{A} \in \{0,1\}^{|\mathcal{U}|\times|\mathcal{V}|}$
B : item-item similarity weight , $\mathbf{B} \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$

$$\underset{\mathbf{B}}{\operatorname{argmin}} \boxed{\|\mathbf{A} - \mathbf{AB}\|_F^2} + \lambda_F \|\mathbf{B}\|_F^2 \quad \text{s.t.} \quad \textsc{Diag}(\mathbf{B}) = 0.$$

* DIAG(): diagonal elements

Aggregator Architecture

$$\underset{\mathbf{B}}{\operatorname{argmin}} \|\mathbf{A} - \mathbf{AB}\|_F^2 + \lambda_F \|\mathbf{B}\|_F^2 \quad \text{s.t.} \quad \textsc{Diag}(\mathbf{B}) = 0.$$

$$\mathbf{B} = \mathbf{I} - \mathbf{P} \cdot \textsc{DiagMat}\left(\mathbf{1} \oslash \textsc{Diag}(\mathbf{P})\right), \quad \text{where } \mathbf{P} = (\mathbf{A}^\top \mathbf{A} + \lambda_F \mathbf{I})^{-1}$$

* DIAGMAT(): diagonal matrix
1 : vector of ones
$\oslash$: element-wise division

Algorithm 1: Item similarity pre-processing in Python

Input: Interaction matrix A and hyperparameter λ_F
Output: Item-item similarity matrix B

```
G = (A.T @ A).toarray() # The item-item coherent matrix
diag_indices = numpy.diag_indices_from(G)
G[diag_indices] += λF   # G + λF * I
P = numpy.linalg.inv(G) # The inverse matrix
B = P / (-numpy.diag(P))
B[diag_indices] = 0
```

Aggregator Architecture

$$\mathbf{B} = \mathbf{I} - \mathbf{P} \cdot \textsc{DiagMat}\left(\mathbf{1} \oslash \textsc{Diag}(\mathbf{P})\right), \quad \text{where } \mathbf{P} = (\mathbf{A}^\top \mathbf{A} + \lambda_F \mathbf{I})^{-1}$$

A =

	v1	v2	v3
u1	1	1	0
u2	1	1	1
u3	0	1	1
u4	0	0	1

(A.T)A =

	v1	v2	v3
v1	2	2	1
v2	2	3	2
v3	1	2	3

Aggregator Architecture

$$\mathbf{B} = \mathbf{I} - \mathbf{P} \cdot \textsc{DiagMat}\big(\mathbf{1} \oslash \textsc{Diag}(\mathbf{P})\big), \quad \text{where } \mathbf{P} = (\mathbf{A}^\top \mathbf{A} + \lambda_F \mathbf{I})^{-1}$$

λF = 0

A =

	v1	v2	v3
u1	1	1	0
u2	1	1	1
u3	0	1	1
u4	0	0	1

(A.T)A =

	v1	v2	v3
v1	2	2	1
v2	2	3	2
v3	1	2	3

P =

	v1	v2	v3
v1	5/3	-4/3	1/3
v2	-4/3	5/3	-2/3
v3	1/3	-2/3	2/3

Aggregator Architecture

$$\mathbf{B} = \mathbf{I} - \mathbf{P} \cdot \textsc{DiagMat}\left(\mathbf{1} \oslash \textsc{Diag}(\mathbf{P})\right), \quad \text{where } \mathbf{P} = \left(\mathbf{A}^\top \mathbf{A} + \lambda_F \mathbf{I}\right)^{-1}$$

λF = 0

P =

	v1	v2	v3
v1	5/3	-4/3	1/3
v2	-4/3	5/3	-2/3
v3	1/3	-2/3	2/3

DIAGMAT(1 ⊘ DIAG(P))=

	v1	v2	v3
v1	3/5	0	0
v2	0	3/5	0
v3	0	0	3/2

B =

	v1	v2	v3
v1	0	0.8	-0.5
v2	0.8	0	1
v3	-0.2	0.4	0

Aggregator Architecture

$$\mathbf{B} = \mathbf{I} - \mathbf{P} \cdot \textsc{DiagMat}\big(\mathbf{1} \oslash \textsc{Diag}(\mathbf{P})\big), \quad \text{where } \mathbf{P} = (\mathbf{A}^\top \mathbf{A} + \lambda_F \mathbf{I})^{-1}$$

λF = 0

A =

	v1	v2	v3
u1	1	1	0
u2	1	1	1
u3	0	1	1
u4	0	0	1

B =

	v1	v2	v3
v1	0	0.8	-0.5
v2	0.8	0	1
v3	-0.2	0.4	0

Aggregator Architecture

- Item-similarity-pooling

$$\boldsymbol{\alpha} = \textsc{Normlize}(\{\alpha_1, \alpha_2, \alpha_3, \ldots\}), \text{ where } \alpha_i = (\mathbf{AB})_{u,v_i}$$

$$\mathbf{h} = \textsc{WeightedMEAN}(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots\}, \boldsymbol{\alpha})\mathbf{W}_{\text{agg}}$$

A =

	v1	v2	v3
u1	1	1	0
u2	1	1	1
u3	0	1	1
u4	0	0	1

AB =

	v1	v2	v3
u1	0.8	0.8	0.5
u2	0.6	1.2	0.5
u3	0.6	0.4	1
u4	-0.2	0.4	1

Aggregator Architecture

- User representation

$$\mathbf{h}_u^X = \lambda_A \cdot \mathbf{u}^X + (1 - \lambda_A) \cdot \textsc{Aggregator}^X(\mathbf{u}^X, \mathcal{H}_u^X),$$
$$\mathbf{h}_u^S = \lambda_A \cdot \mathbf{u}^S + (1 - \lambda_A) \cdot \textsc{Aggregator}^S(\mathbf{u}^S, \mathcal{H}_u^S),$$

Loss function

- Contrastive loss

positive negative

$$\mathcal{L}_{\text{con}}^{X} = \sum_{u \in \mathcal{U}^X} \left[-\log \textsc{Disc}^X(\widetilde{\mathbf{h}}_u^X, \widehat{\mathbf{h}}_{u,X}^S) - \log(1 - \textsc{Disc}^X(\widetilde{\mathbf{h}}_{\bar{u}}^X, \widehat{\mathbf{h}}_{u,X}^S)) \right]$$

$$\boxed{\bar{u} \neq u}$$

$$\textsc{Disc}^X(\widetilde{\mathbf{h}}_u^X, \widehat{\mathbf{h}}_{u,X}^S) = \textsc{Sigmoid}\left(\widetilde{\mathbf{h}}_u^X \mathbf{W}_{\text{disc}}^X (\widehat{\mathbf{h}}_{u,X}^S)^\top\right)$$

Interaction masking: $\widetilde{\mathcal{H}}_u = \textsc{Mask}(\mathcal{H}_u, \boldsymbol{m}),$ where $m_i = \textsc{Bernoulli}(p)$

Domain masking: $\widehat{\mathcal{H}}_{u,X}^S = \{\varnothing, \widetilde{\mathcal{H}}_u^Y\}, \quad \widehat{\mathcal{H}}_{u,Y}^S = \{\widetilde{\mathcal{H}}_u^X, \varnothing\}$

Loss function

- Prediction loss

$$\mathcal{L}^{X}_{\text{pred}} = \sum_{(u,v_i)\in\mathcal{E}^X} [-\log y^X_{(u,v_i)} - \log(1 - y^X_{(u,\bar{v}_i)})],$$

$$y^X_{(u,v_i)} = \textsc{Sigmoid}\big(\textsc{Score}^X(\widetilde{\mathbf{h}}^X_u, \mathbf{v}^X_i) + \textsc{Score}^X(\widehat{\mathbf{h}}^S_{u,X}, \mathbf{v}^X_i)\big), \quad \text{score} \in \{0, 1\}$$

- Training loss

$$\mathcal{L} = \lambda(\mathcal{L}^X_{\text{pred}} + \mathcal{L}^Y_{\text{pred}}) + (1-\lambda)(\mathcal{L}^X_{\text{con}} + \mathcal{L}^Y_{\text{con}}).$$

Outline

- Introduction
- Method
- **Experiment**
- Conclusion

Experiment

- **Scenario**

Scenarios	Domains		Overlapping		Recommendation	
	Dual	Multi	User	Item	Intra	Inter
Scenario 1	✔		✔		✔	
Scenario 2	✔		✔			✔

- **Evaluation**
  - Hit Ratio(HR)
  - Normalized Discounted Cumulative Gain(NDCG)

Experiment

- **Dataset**

Use users with number of ratings greater than 5 and items with number of ratings greater than 10 in the **Amazon** dataset.

  - Scenario 1 : Sport & Cloth, Elec & Phone (only use the overlapping user)
  - Scenario 2 : Sport & Cloth, Game & Video

- **Evaluation**
  - Hit Ratio(HR)
  - Normalized Discounted Cumulative Gain(NDCG)

Experiment

Scenario 1 (user-attention)

MLP-based | GNN-based

Datasets	Metric@10	Cross-Domain Methods					Ours
		CoNet	DDTCDR	PPGN	Bi-TGCF	DisenCDR	UniCDR
Sport	HR	12.09	11.86	15.10	14.83	17.55	**18.37**
	NDCG	6.41	6.37	8.03	7.95	9.46	**10.98**
Cloth	HR	12.40	12.54	14.23	14.68	16.31	**17.85**
	NDCG	6.62	7.13	7.68	7.93	9.03	**11.20**
Elec	HR	17.22	18.47	21.68	22.14	**24.57**	22.92
	NDCG	9.86	11.08	11.63	12.20	**14.51**	13.83
Phone	HR	17.66	17.23	24.54	25.71	**28.76**	24.72
	NDCG	9.30	8.58	13.34	13.93	**16.13**	13.77

Scenario 2 (mean)

embedding and mapping

Datasets	Metric@10	Cross-Domain Methods					Ours
		EMCDR	SSCDR(CML)	TMCDR	SA-VAE	CDRIB	UniCDR
Sport	HR	7.44	7.27	7.18	7.51	**12.04**	11.20
	NDCG	3.71	3.75	3.84	3.72	6.22	**7.04**
Cloth	HR	7.29	6.12	8.11	7.21	12.19	**12.48**
	NDCG	4.48	3.06	5.05	4.59	6.81	**7.52**
Game	HR	4.63	3.48	5.36	5.84	8.51	**8.78**
	NDCG	2.24	1.59	2.58	2.78	4.58	**4.63**
Video	HR	7.94	5.51	8.85	7.46	**13.17**	10.74
	NDCG	4.29	2.61	4.41	3.71	**6.49**	5.89

Experiment

- An appropriate aggregator is needed for different CDR scenarios.

Scenarios	Datasets	Metrics@10	Aggregators		
			Mean	User-Att	Item-Sim
Scenario 1	Sport	HR	14.26	18.37	16.96
		NDCG	8.55	10.98	9.83
	Cloth	HR	15.23	17.85	16.32
		NDCG	9.43	11.20	10.18
	Elec	HR	22.96	23.08	22.92
		NDCG	13.45	13.73	13.83
	Phone	HR	22.86	22.38	24.72
		NDCG	12.64	12.42	13.77
Scenario 2	Sport	HR	11.20	7.57	9.88
		NDCG	7.04	4.36	5.95
	Cloth	HR	12.48	9.25	10.89
		NDCG	7.52	5.79	6.46
	Game	HR	8.78	6.37	7.44
		NDCG	4.63	3.33	3.82
	Video	HR	10.74	5.71	10.38
		NDCG	5.89	3.12	5.67

Experiment

- Domain information play an essential role in the inter-domain recommendation.

Scenarios	Datasets	Metrics@10	Mask Mechanism		
			w/o Inter	w/o Domain	UniCDR
Scenario 1	Sport	HR	17.03	18.55	18.37
		NDCG	10.30	11.04	10.98
	Cloth	HR	17.21	17.15	17.85
		NDCG	10.26	10.93	11.20
	Elec	HR	22.18	22.38	22.92
		NDCG	11.15	13.80	13.83
	Phone	HR	22.74	25.86	24.72
		NDCG	12.45	14.38	13.77
Scenario 2	Sport	HR	10.36	7.50	11.20
		NDCG	6.33	4.38	7.04
	Cloth	HR	11.97	7.71	12.48
		NDCG	7.45	4.29	7.52
	Game	HR	7.96	6.80	8.78
		NDCG	4.01	3.21	4.63
	Video	HR	9.87	6.55	10.74
		NDCG	5.02	3.29	5.89

* Inter: interaction mask
Domain : domain mask

Conclusion

- Well-learned domain-shared information can help both inter-domain and intra-domain recommendations.