Beyond Co-occurrence:Multi-modal Session-based Recommendation

Xiaokun Zhang, Bo Xu, Fenglong Ma, Chenliang Li, Liang Yang and Hongfei Lin

Xiaokun Zhang is currently pursuing the PhD degree with the School of Computer Science and T echnology, Dalian University of Technology,China.

Bo Xu received the BSc and PhD degrees from the Dalian University of T echnology, China, in 2011 and 2018.

Fenglong Ma is an assistant professor in the College of Information Sciences and Technology at the Pennsylvania State University.

Chenliang Li is a full Professor with School of Cyber Science and Engineering, Wuhan University China.

Liang Yang received the BSc and PhD degrees from the Dalian University of T echnology, China, in 2009 and 2017, respectively.

Hongfei Lin received the BSc degree from the Northeastern Normal University in 1983, the MSc degree from the Dalian University of Technology in 1992, and the PhD degree from the Northeastern University in 2000.

code: https://github.com/Zhang-xiaokun/MMSBR.

TKDE 2023

Reported by Minqin Li

Introduction

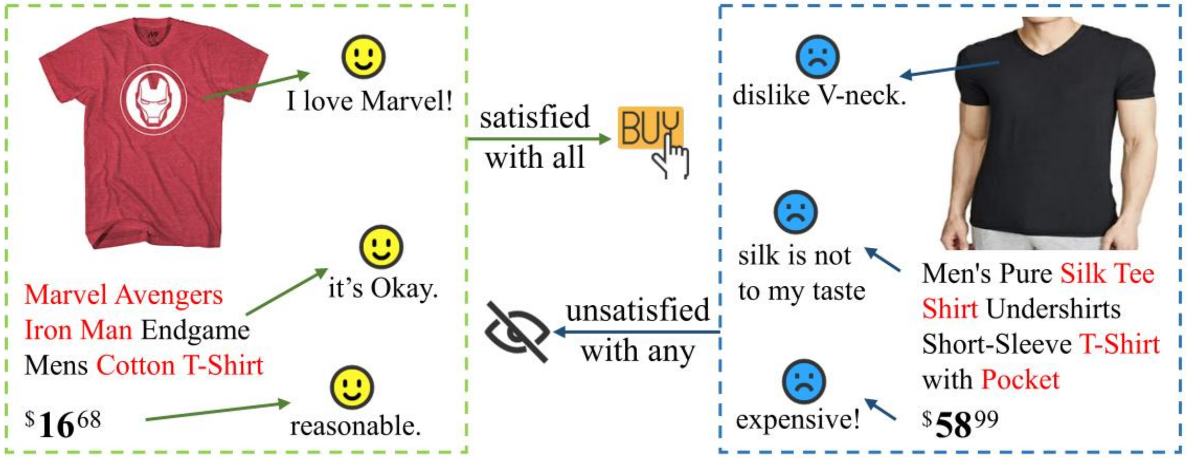


Fig. 1: A user makes the decision after evaluating all multi-modal information displayed on pages including item images, description text and price.

Existing methods mostly focus on mining limited item co-occurrence patterns exposed by item ID within sessions, while ignoring what attracts users to engage with certain items is rich multi-modal information displayed on pages.

(1) How to extract relevant semantics from heterogeneous descriptive information with different noise?

(2)How to fuse these heterogeneous descriptive information to comprehensively infer user interests?

(3)How to handle probabilistic influence of numerical information on user behaviors?

Method

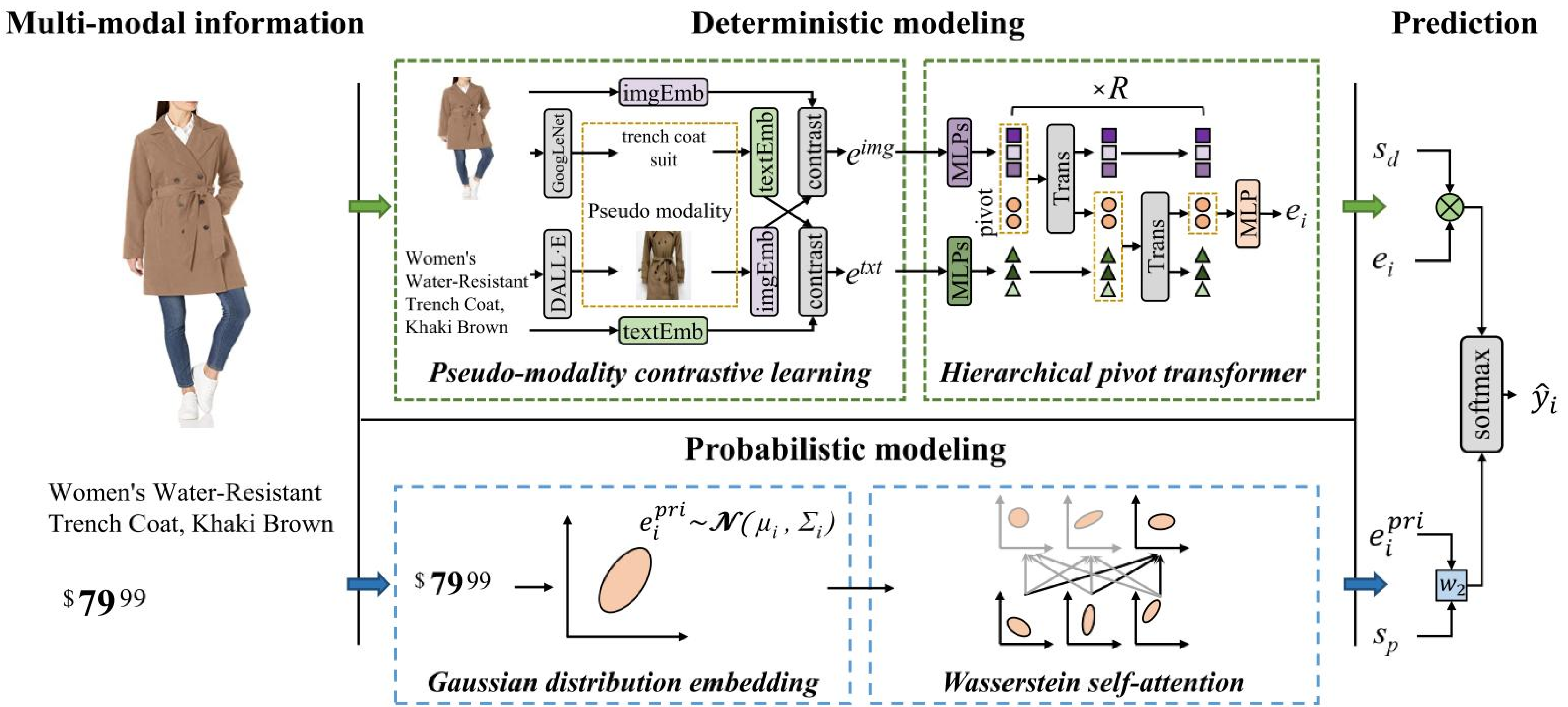
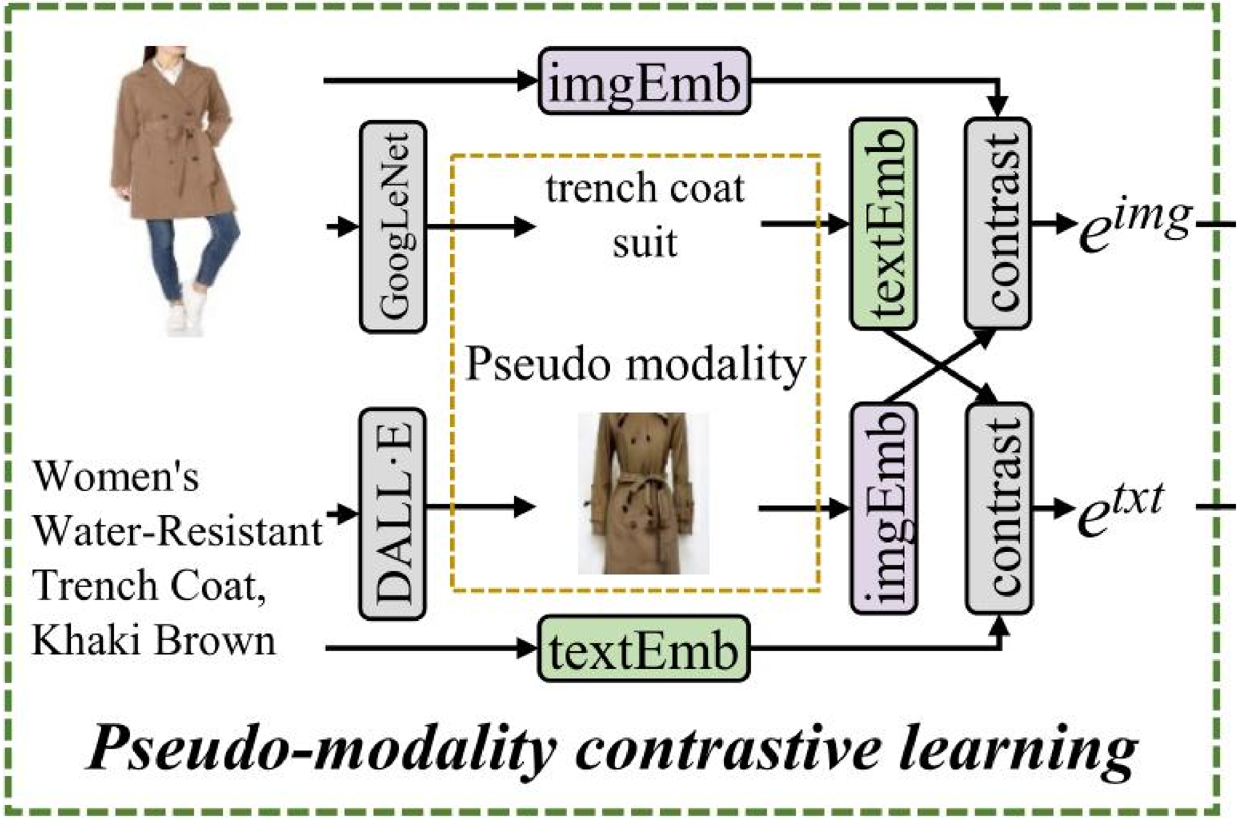


Fig. 2:The proposed MMSBR customizes deterministic and probabilistic modeling to handle descriptive and numerical information respectively.

Method



$$\mathbf{e}_i^{img} = \mathrm{imgEmb}(x_i^{img}) \tag{1}$$

$$\mathbf{e}_i^{txt} = \mathrm{textEmb}(x_i^{txt}) \tag{2}$$

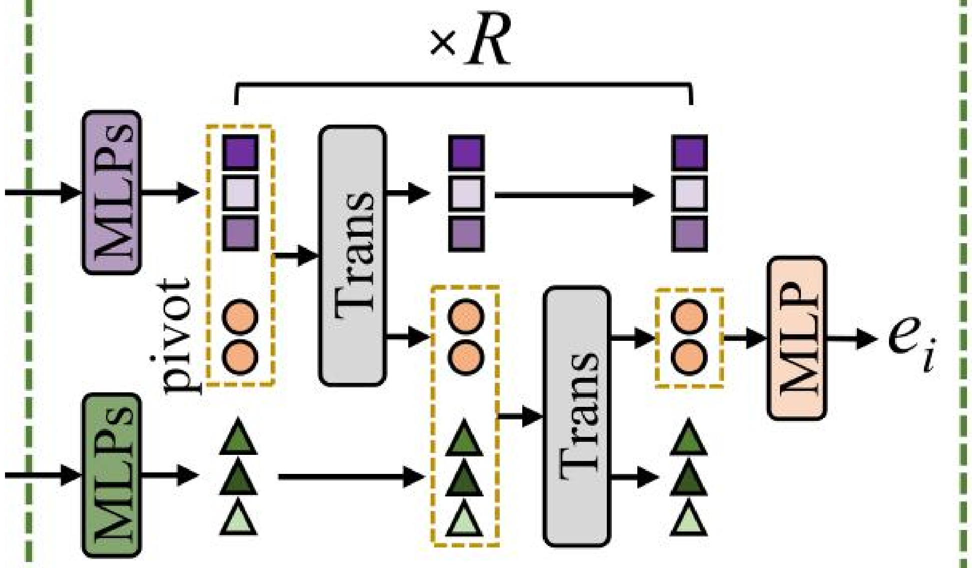
$$v_i^{pri} = \lfloor \frac{x_i^{pri} - \min}{\max - \min} \times \rho \rfloor \tag{3}$$

$$\mathbf{e}_i^{pseimg} = \mathrm{imgEmb}(x_i^{pseimg}) \tag{4}$$

$$\mathbf{e}_i^{psetxt} = \mathrm{textEmb}(x_i^{psetxt}) \tag{5}$$

$$\mathcal{L}_{con} = -\frac{\exp(\mathrm{sim}(\mathbf{e}_i^{img}, \mathbf{e}_i^{pseimg}))}{\sum_{k=1}^{n} \exp(\mathrm{sim}(\mathbf{e}_i^{img}, \mathbf{e}_k^{pseimg}))} - \frac{\exp(\mathrm{sim}(\mathbf{e}_i^{txt}, \mathbf{e}_i^{psetxt}))}{\sum_{k=1}^{n} \exp(\mathrm{sim}(\mathbf{e}_i^{txt}, \mathbf{e}_k^{psetxt}))}, \tag{6}$$

Method



Hierarchical pivot transformer

$$\mathbf{Z}_{img} = \{\mathrm{MLP}_1^{img}(\mathbf{e}_i^{img}), \ldots, \mathrm{MLP}_C^{img}(\mathbf{e}_i^{img})\} \quad (7)$$

$$\mathbf{Z}_{txt} = \{\mathrm{MLP}_1^{txt}(\mathbf{e}_i^{txt}), \ldots, \mathrm{MLP}_C^{txt}(\mathbf{e}_i^{txt})\} \quad (8)$$

$$\mathbf{F}_*^l = \mathrm{MSA}(\mathrm{LN}(\mathbf{F}^l)) + \mathbf{F}^l \quad (9)$$

$$\mathbf{F}^{l+1} = \mathrm{FCL}(\mathrm{LN}(\mathbf{F}_*^l)) + \mathbf{F}_*^l \quad (10)$$

$$[\mathbf{Z}_{img}^{l+1}, \mathbf{P}_{img}^l] = \mathrm{Trans}([\mathbf{Z}_{img}^l, \mathbf{P}^l]) \quad (11)$$

$$\mathbf{p}_*^l = (\mathbf{P}_{img}^l + \mathbf{P}^l)/2 \quad (12)$$

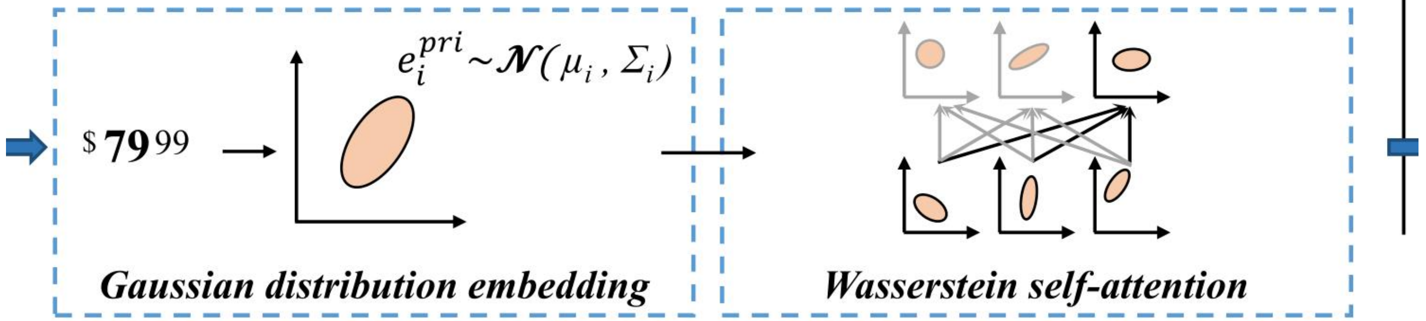
$$[\mathbf{Z}_{txt}^{l+1}, \mathbf{P}_{txt}^l] = \mathrm{Trans}([\mathbf{Z}_{txt}^l, \mathbf{P}_*^l]) \quad (13)$$

$$\mathbf{p}^{l+1} = (\mathbf{P}_{txt}^l + \mathbf{P}_*^l)/2 \quad (14)$$

$$\mathbf{e}_i = \mathrm{MLP}(\mathbf{P}^R) = \mathrm{MLP}([\mathbf{p}_1^R; \mathbf{p}_2^R; \ldots; \mathbf{p}_T^R]) \quad (15)$$

Method

Probabilistic modeling



$$\mathbf{s}_d = \sum_{k=1}^{m} \alpha_k \mathbf{e}_k \tag{16}$$

$$\alpha_k = \mathbf{u}\sigma(\mathbf{A}_1\mathbf{e}_k + \mathbf{A}_2\bar{\mathbf{e}} + \mathbf{b}) \tag{17}$$

$$\hat{\mathbf{e}}_i^{pri} = \text{Gaussian}(v_i^{pri}) \sim \mathcal{N}(\hat{\mu}_i, \hat{\Sigma}_i) \tag{18}$$

$$\mathbf{e}_i^{pri} \sim \mathcal{N}(\mu_i, \Sigma_i) = \mathcal{N}(\hat{\mu}_i + \mathbf{e}_i^c, \hat{\Sigma}_i + \mathbf{e}_i^c) \tag{19}$$

$$\mathcal{W}_2(\mathcal{G}_1, \mathcal{G}_2) = \sqrt{\|\mu_1 - \mu_2\|_2^2 + \left\|(\mathbf{\Sigma}_1)^{\frac{1}{2}} - (\mathbf{\Sigma}_2)^{\frac{1}{2}}\right\|_2^2} \tag{20}$$

$$\mathbf{H} = \text{WSA}(A^Q\mathbf{E}_p, A^K\mathbf{E}_p, A^V\mathbf{E}_p) \tag{21}$$

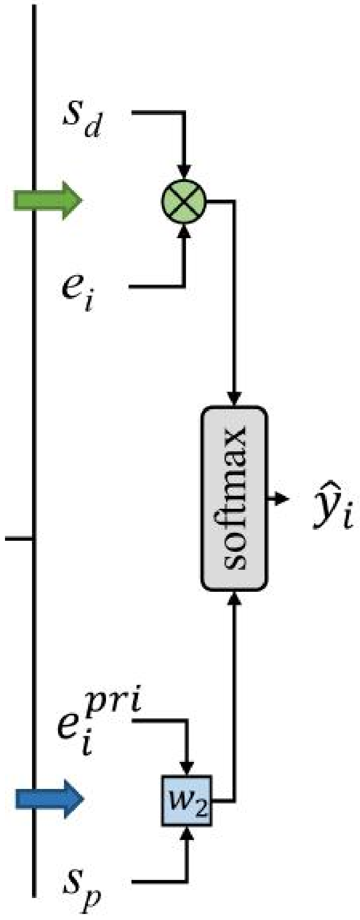
$$\mathbf{h}_i^{\mu} = \sum_{j=1}^{m} a_{ij} A_{\mu}^V \mu_j, \text{ and } \mathbf{h}_i^{\Sigma} = \sum_{j=1}^{m} a_{ij}^2 A_{\Sigma}^V \mathbf{\Sigma}_j \tag{22}$$

$$\mathbf{h}_i^{\mu} = \sum_{j=1}^{m} a_{ij} A_{\mu}^V \mu_j, \text{ and } \mathbf{h}_i^{\Sigma} = \sum_{j=1}^{m} a_{ij}^2 A_{\Sigma}^V \mathbf{\Sigma}_j \tag{23}$$

$$\mathbf{s}_p = \mathbf{h}_m \sim \mathcal{N}(\mathbf{h}_m^{\mu}, \mathbf{h}_m^{\Sigma}) \tag{24}$$

Method

Prediction



$$\hat{y}_i = softmax(\mathbf{e}_i \mathbf{s}_d + \mathcal{W}_2(\mathbf{e}_i^{pri}, \mathbf{s}_p)) \tag{25}$$

$$\mathcal{L}_{rec} = -\sum_{i=1}^{n} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \tag{26}$$

$$\mathcal{L} = \mathcal{L}_{rec} + \lambda \mathcal{L}_{con} \tag{27}$$

Experiments

TABLE 2: Statistics of all datasets.

Datasets	Cellphones	Grocery	Sports
#item	8,614	11,638	18,796
#category	48	665	1,259
#interaction	196,376	364,728	566,504
#session	78,026	127,548	211,959
avg.length	2.52	2.86	2.67

TABLE 5: Statistics of datasets with cold-start items.

Datasets	Cellphones+	Grocery+	Sports+
#item	10,245(+1631)	13,493(+1855)	22,049(+3253)
#category	48(-)	678(+13)	1,312(+53)
#interaction	199,065(+2689)	367,674(+2946)	571,789(+5285)
#session	78,987(+961)	128,510(+962)	213,787(+1828)
avg.length	2.52(-)	2.86(-)	2.67(-)

Experiments

TABLE 3: Performance comparison of MMSBR with baselines over three datasets. The results (%) produced by the best baseline and the best performer in each column are underlined and boldfaced respectively. Statistical significance of pairwise differences for MMSBR against the best baseline (*) is determined by the t-test ($p < 0.01$).

Method	Cellphones				Grocery				Sports			
	Prec@10	MRR@10	Prec@20	MRR@20	Prec@10	MRR@10	Prec@20	MRR@20	Prec@10	MRR@10	Prec@20	MRR@20
S-POP	5.32	2.71	7.24	2.85	20.65	17.00	23.64	17.25	15.61	14.56	17.59	14.69
SKNN	21.07	9.95	24.71	10.21	39.83	25.15	41.88	25.29	31.79	21.31	33.86	21.46
NARM	20.59	15.32	24.12	15.56	40.39	34.53	42.41	34.62	31.64	26.94	34.17	27.12
SASRec	23.37	15.47	27.58	15.76	40.97	34.76	43.02	34.92	31.54	26.68	34.11	26.87
BERT4Rec	22.28	14.39	27.09	14.73	40.59	34.09	42.93	34.31	31.57	26.85	34.32	27.07
SR-GNN	21.80	15.60	25.08	15.77	40.81	34.89	42.74	35.01	31.96	27.43	34.29	27.51
COTREC	23.78	10.82	28.33	11.13	41.28	30.60	43.24	30.75	32.16	23.28	35.13	23.46
MSGIFSR	20.92	14.53	24.51	14.77	41.34	35.25	43.40	35.47	32.28	27.56	34.95	27.72
MGS	21.74	14.29	25.21	14.54	40.92	35.06	42.79	35.20	31.63	26.75	33.76	26.89
UniSRec	22.73	15.36	26.65	15.63	41.40	35.12	43.44	35.24	31.90	26.91	34.41	27.04
CoHHN	23.60	15.77	27.71	15.96	41.58	35.33	43.59	35.58	32.12	27.13	35.02	27.31
MMSBR	**24.37***	**16.47***	**29.22***	**16.81***	**42.10***	**35.91***	**44.27***	**36.06***	**32.89***	**28.10***	**35.64***	**28.28***

Experiments

TABLE 4: The effect of hierarchical pivot transformer.

Method	Cellphones		Grocery		Sports	
	Prec@20	MRR@20	Prec@20	MRR@20	Prec@20	MRR@20
COTREC	28.33	11.13	43.24	30.75	35.13	23.46
MSGIFSR	24.51	14.77	43.40	35.47	34.95	27.72
MMSBR_{mlp}	26.74	15.95	42.93	35.28	34.67	27.86
MMSBR	**29.22***	**16.81***	**44.27***	**36.06***	**35.64***	**28.28***

Experiments

TABLE 6: The influence of different modalities.

Method	Cellphones		Grocery		Sports	
	Prec@20	MRR@20	Prec@20	MRR@20	Prec@20	MRR@20
(a) w/o image	27.45	14.85	41.23	35.20	32.14	27.50
(b) w/o text	27.19	14.69	41.11	35.08	32.22	27.42
(c) w/o price	25.10	13.35	42.98	35.57	34.78	27.68
MMSBR	**29.22***	**16.81***	**44.27***	**36.06***	**35.64***	**28.28***

Experiments

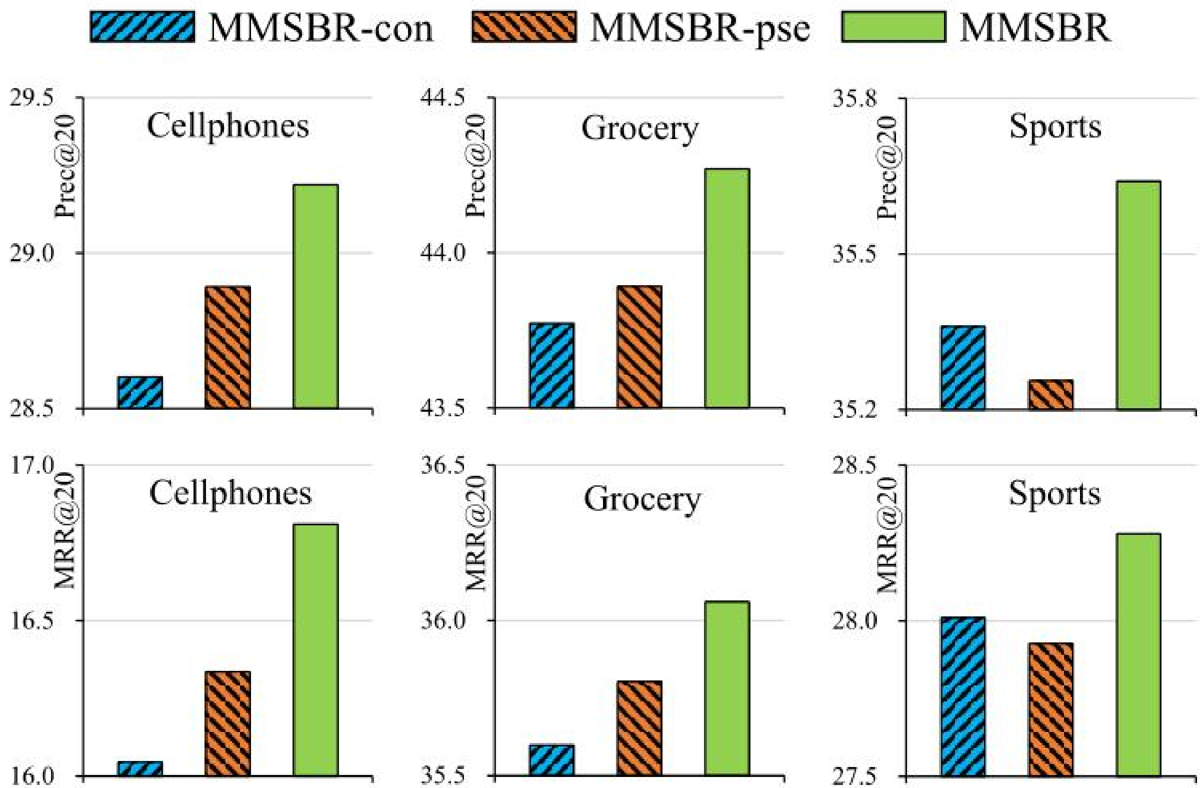


Fig. 3: The effect of pseudo-modality contrastive learning.

Experiments

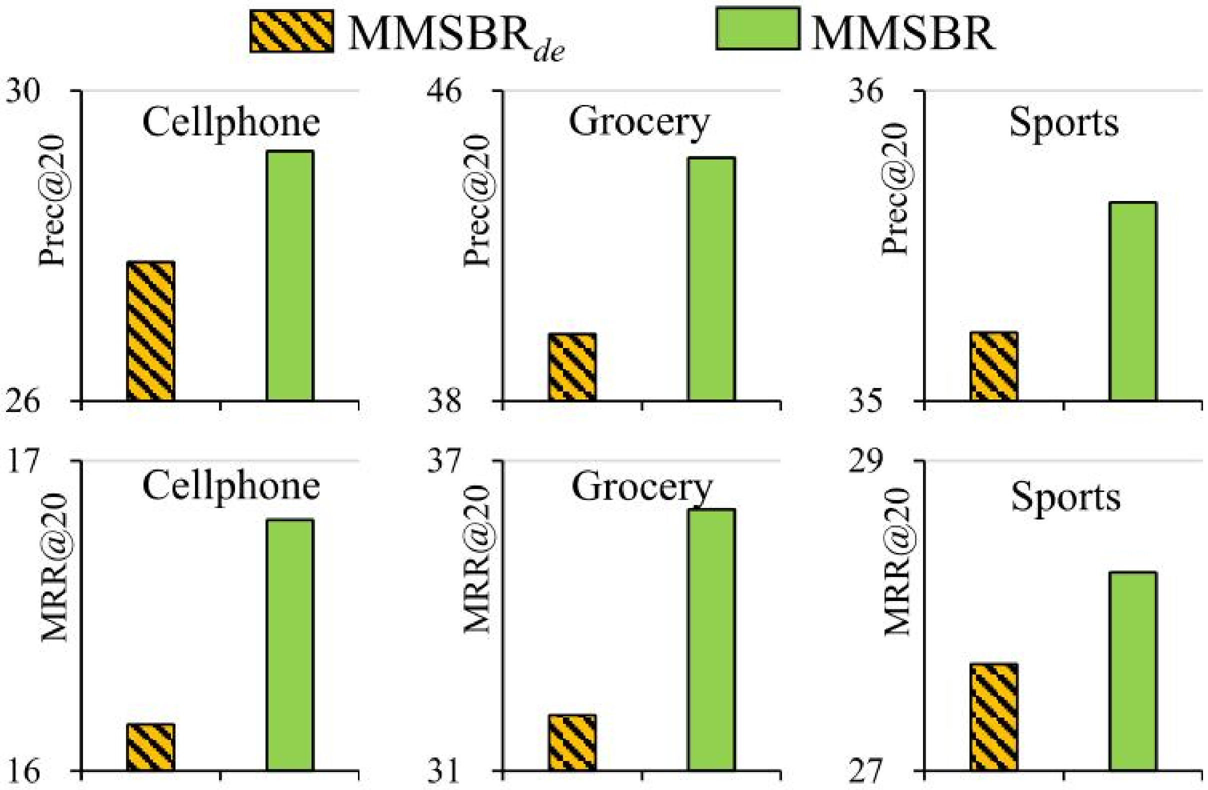


Fig. 4: The effect of probabilistic modeling.

Experiments

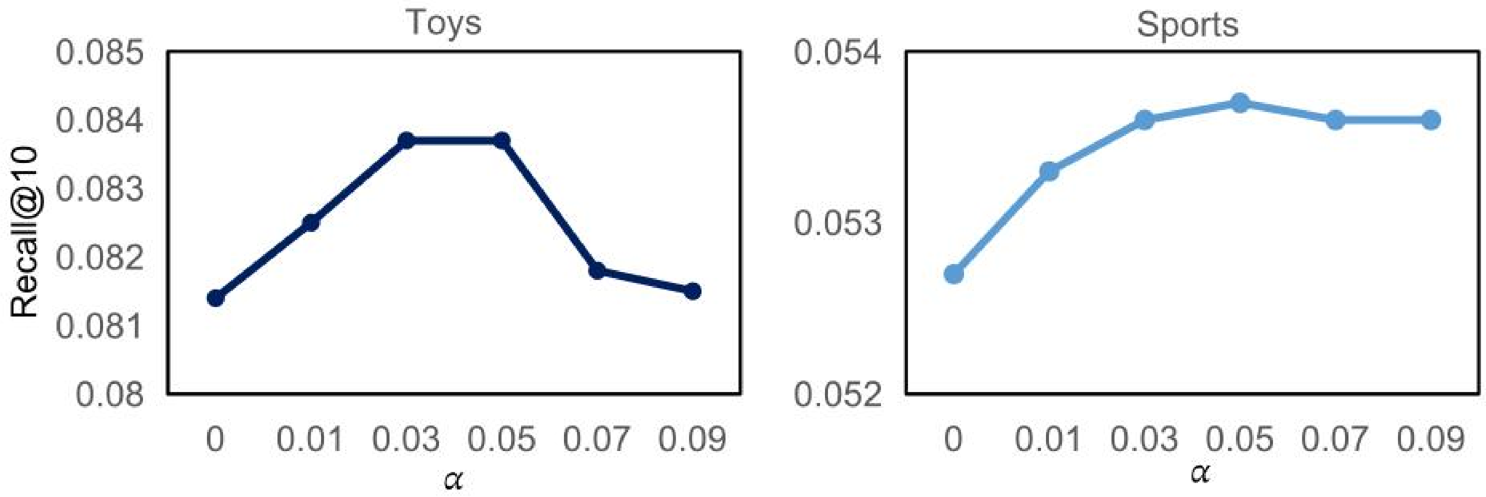


Figure 4: Effect of balance parameter α.

Experiments

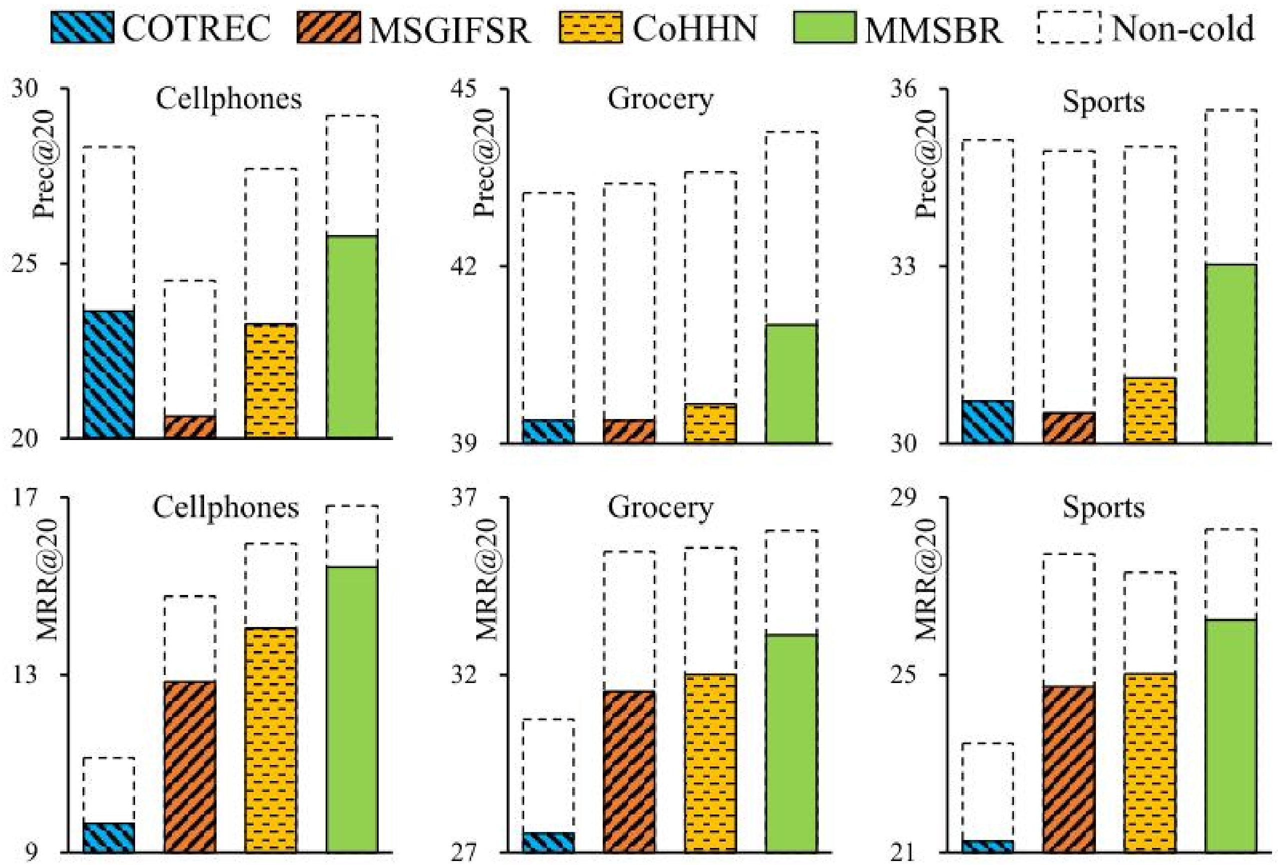


Fig. 5: Performance in cold-start scenario.

Experiments

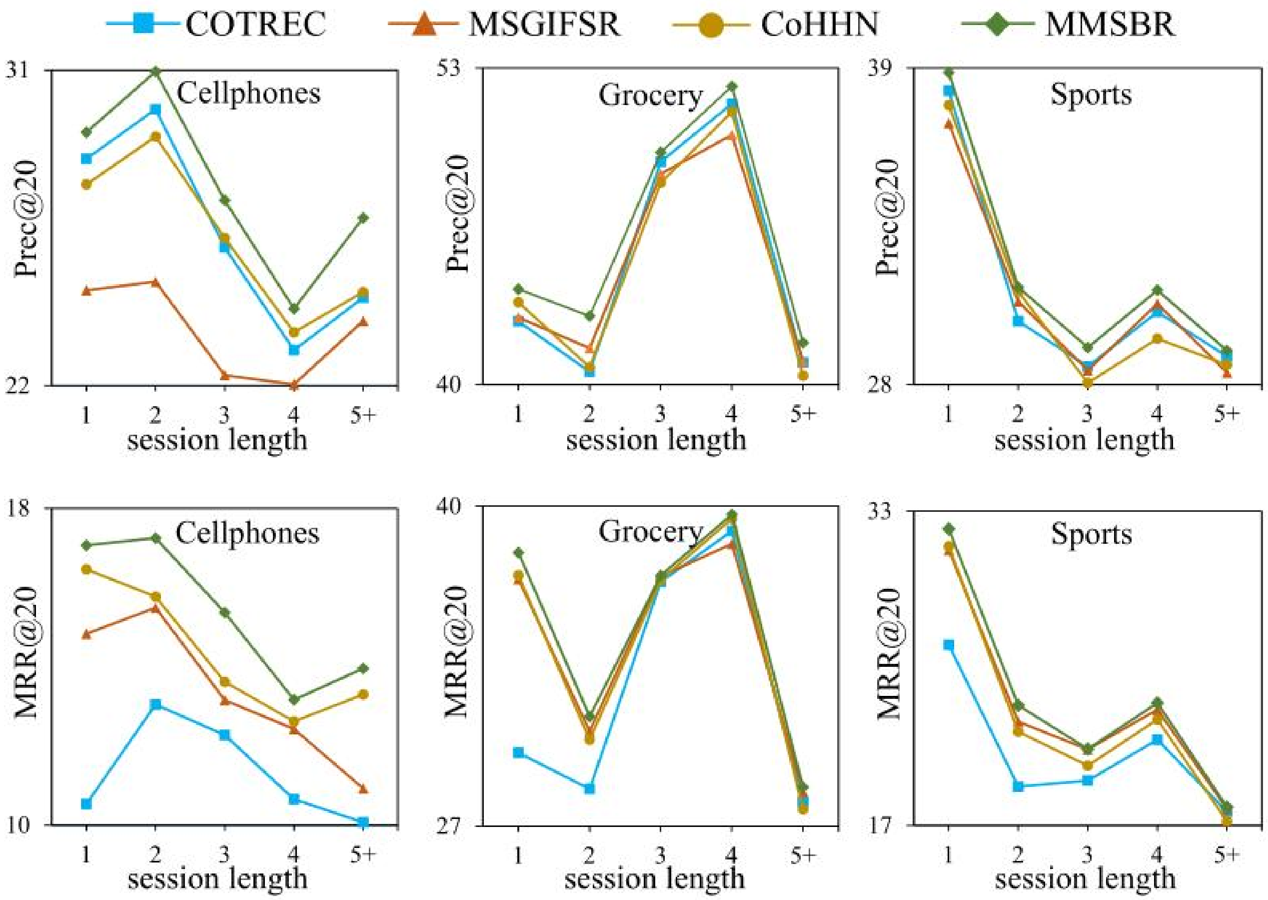


Fig. 6: Impact of various session lengths.

Experiments

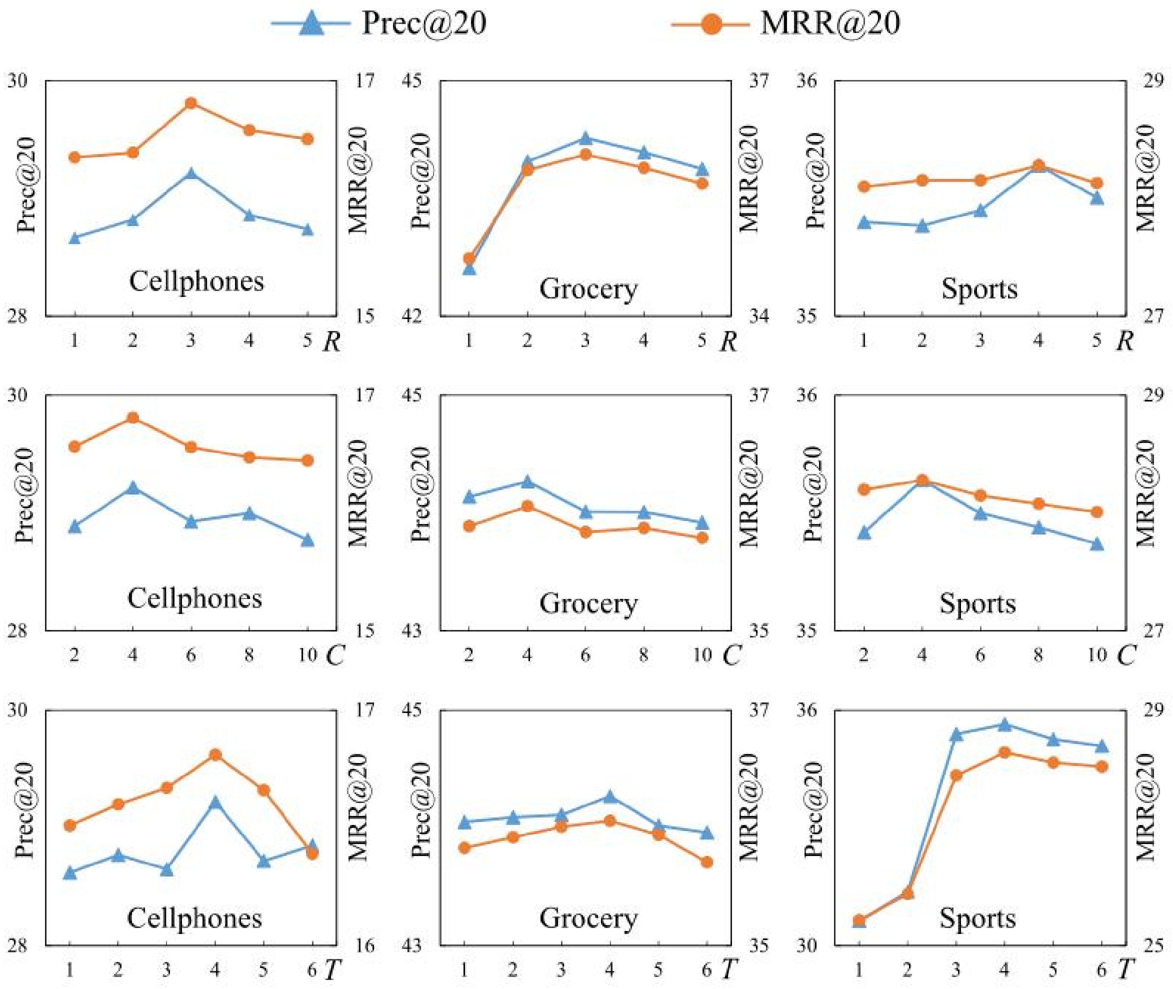


Fig. 7: Impact of hyperparameters.

Thanks