Contrastive Meta Learning with Behavior Multiplicity for Recommendation（WSDM_2022）

Wei Wei , Chao Huang , Lianghao Xia, Yong Xu, Jiashu Zhao, Dawei Yin
Department of Computer Science, Musketeers Foundation Institute of Data Science, University of Hong Kong
South China University of Technology, Wilfrid Laurier University, Baidu Inc
weiwei1206cs@gmail.com, chaohuang75@gmail.com, aka_xia@foxmail.com,
yxu@scut.edu.cn, jzhao@wlu.ca, yindawei@acm.org

Code&data : https://github.com/weiwei1206/CML

2022. 6. 22 • ChongQing

gesis Leibniz-Institut für Sozialwissenschaften

Forschungszentrum · Research Center L3S

Reported by Lele Duan

1.Background

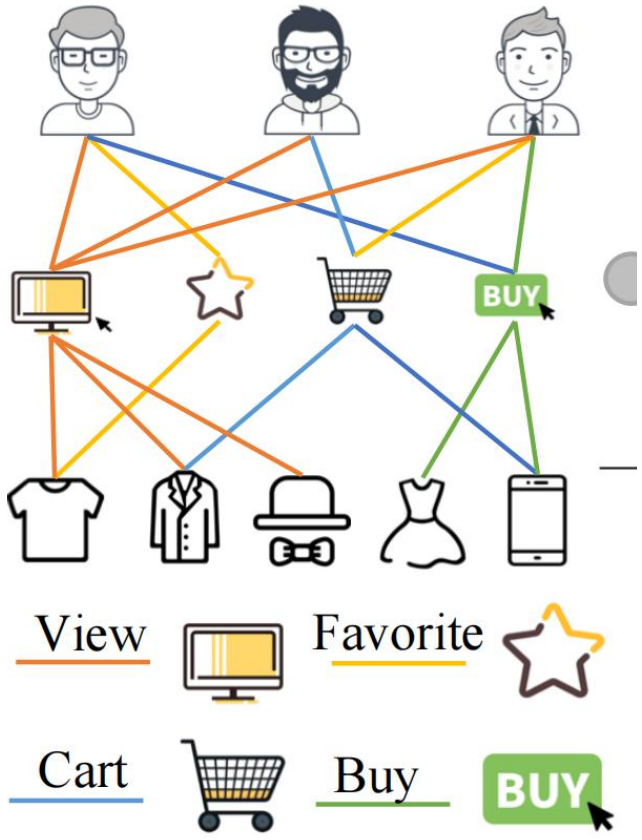
2.Method

3.Experiments

Background

- *Fail to deal with the sparse supervision signal under target behaviors.*
  - ➢ *Multi-behavior contrastive learning framework to distill transferable knowledge across different types of behaviors via the constructed contrastive loss*

- *Fail to capture the personalized multi-behavior patterns with customized dependency modeling.*
  - ➢ *Contrastive meta network to encode the customized behavior heterogeneity for different users*



➢ Over view

Input: *observed user-item interactions with multiplex K types $\{\mathcal{X}^{(1)},...,\mathcal{X}^{(k)},...,\mathcal{X}^{(K)}\}$ among users $\mathcal{U}$ and items $\mathcal{I}$.*

Input: *a predictive function which estimates the likelihood of user $\mathcal{U}$ will interact with item $\mathcal{I}$ under the target type (k) of behaviors.*

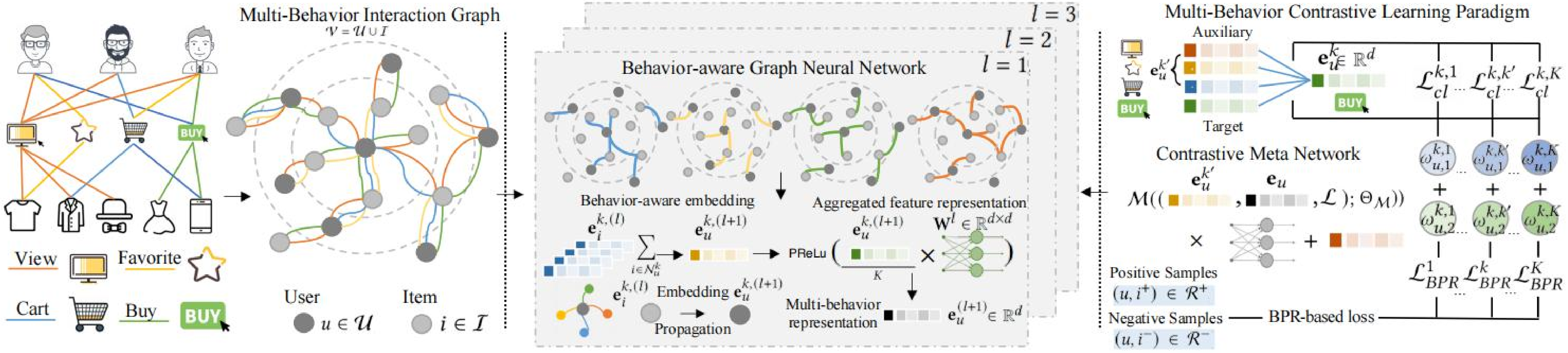
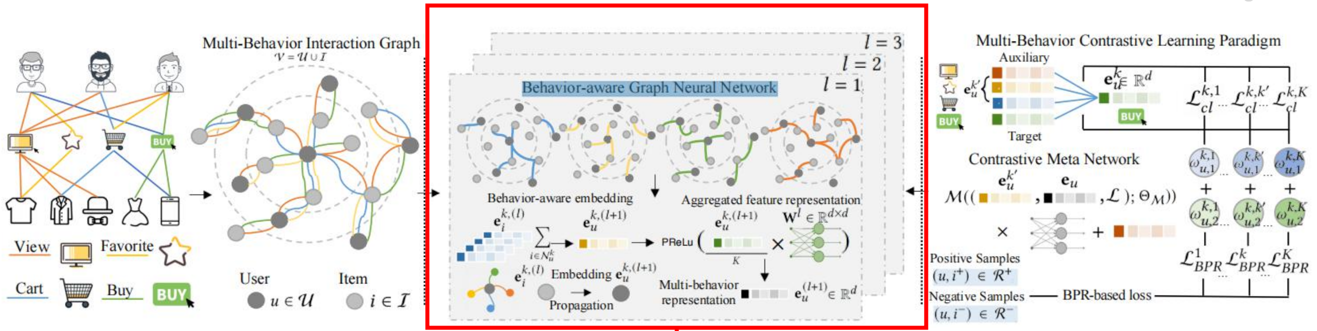


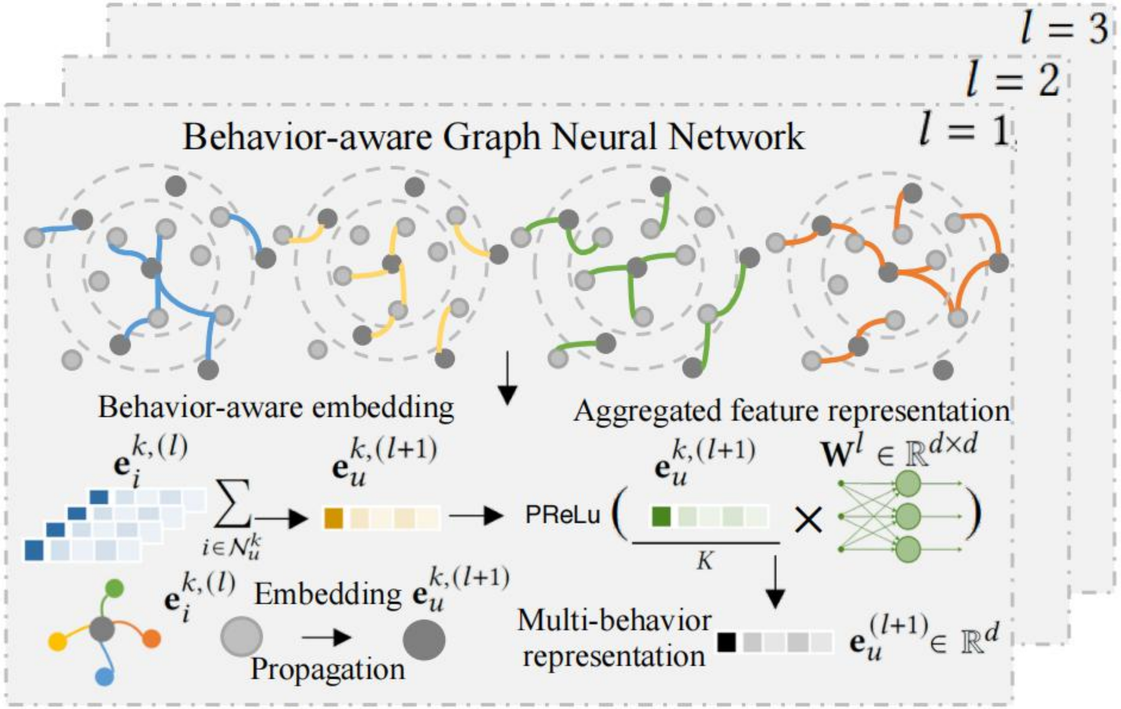
Figure 1: The model flow of CML framework. i) The designed graph neural network $\mathcal{G}(\mathcal{A};\Theta_{\mathcal{G}})$ performs the behavior-aware message passing over the multi-behavior interaction graph $G = \{\mathcal{V},\mathcal{E}\}$. ii) The contrastive views are constrcuted between auxiliary and target behavior embeddings $\mathbf{e}_u^k$, $\mathbf{e}_u^{k'}$. iii) Our proposed meta contrastive encoder captures the customized cross-type behavior dependency with the meta weight network $\mathcal{M}((\mathcal{L},\mathbf{E},\mathbf{E}^k);\Theta_{\mathcal{M}})$. $\omega_u^{k,k'}$ is the personalized contrastive loss weight.

Behavior-aware Graph Neural Network



$$\mathbf{e}_u^{k,(l+1)} = \sum_{i \in \mathcal{N}_u^k} \mathbf{e}_i^{k,(l)}; \quad \mathbf{e}_i^{k,(l+1)} = \sum_{u \in \mathcal{N}_i^k} \mathbf{e}_u^{k,(l)} \tag{1}$$

where $\mathbf{e}_v^{k,(l+1)} \in \mathbb{R}^d$ is defined as the obtained representation of node v ($v \in \{u, i\}$) under the l-th graph neural layer. $\mathcal{N}_u^k$ and $\mathcal{N}_i^k$ denotes the neighboring nodes of item i and user u, respectively.

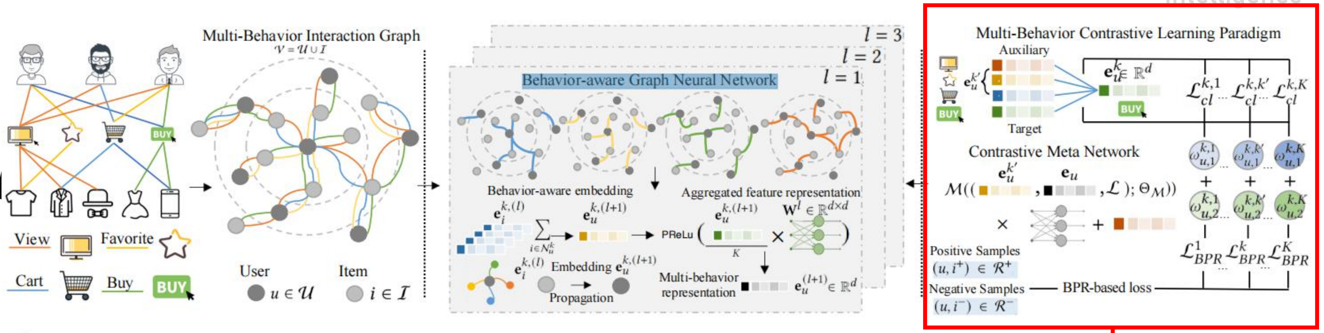


$$\mathbf{e}_u^{(l+1)} = \mathrm{PReLu}\Big(\mathbf{W}^l \cdot \frac{\sum_{k \in K} \mathbf{e}_u^{k,(l+1)}}{K}\Big) \tag{2}$$

The aggregated feature representation $\mathbf{e}_u^{(l+1)}$ could preserve multi-behavior contextual information. $\mathbf{W}^l \in \mathbb{R}^{d \times d}$ represents the transformation matrix corresponding to l-th graph propagation layer.

Multi-Behavior Contrastive learning

- Behavior-Wise Contrastive Learning Paradigm

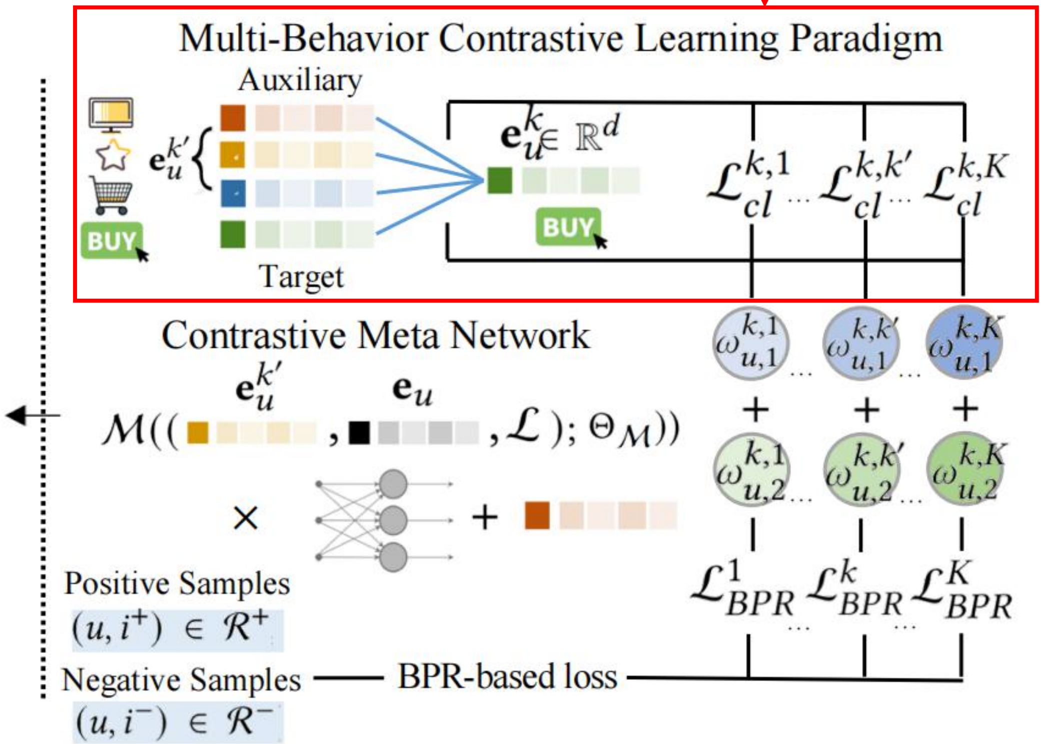


pairs. Given the encoded target behavior representation $\mathbf{e}_u^k$ from our graph neural architecture, the generated positive and negative pairs are $\{\mathbf{e}_u^k, \mathbf{e}_u^{k'} | u \in \mathcal{U}\}$ and $\{\mathbf{e}_u^k, \mathbf{e}_{u'}^{k'} | u, u' \in \mathcal{U}, u \neq u'\}$. The incorpo-

$$\mathcal{L}_{cl}^{k,k'} = \sum_{u \in \mathcal{U}} -log \frac{\exp(\varphi(\mathbf{e}_u^k, \mathbf{e}_u^{k'})/\tau)}{\sum_{u' \in \mathcal{U}} \exp(\varphi(\mathbf{e}_u^k, \mathbf{e}_{u'}^{k'})/\tau)} \quad (3)$$

Here, we define $\varphi(\cdot)$ as the similarity function (*e.g.*, inner-product or cosine similarity) between two embeddings. τ represents the temperature hyperparameter for the softmax function. To sum up,

$$\mathcal{L}_{cl} = \mathcal{L}_{cl}^{k,1} + ... + \mathcal{L}_{cl}^{k,k'} + ... + \mathcal{L}_{cl}^{k,K}.$$



Method

Meta Contrastive Encoding

- *Meta-Knowledge Encoder*

$$\mathbf{Z}_{u,1}^{k,k'} = (d(\mathcal{L}_{cl}^{k,k'}) \cdot \gamma) \parallel \mathbf{e}_u^{k'} \parallel \mathbf{e}_u; \quad \mathbf{Z}_{u,2}^{k,k'} = \mathcal{L}_{cl}^{k,k'} \cdot (\mathbf{e}_u^{k'} \parallel \mathbf{e}_u) \quad (4)$$

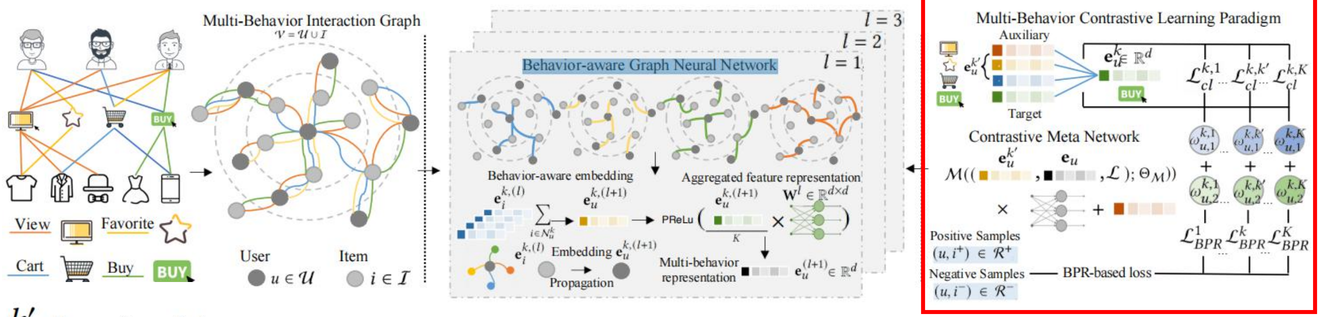
where the encoded meta-knowledge is represented by $\mathbf{Z}_{u,1}^{k,k'}$ and $\mathbf{Z}_{u,2}^{k,k'}$. We define $d(\cdot)$ as the duplicate function to generate a value vector corresponding to the embedding dimensionality. $\parallel$ denotes the concatenation operation. γ is a scale factor for the enlarge

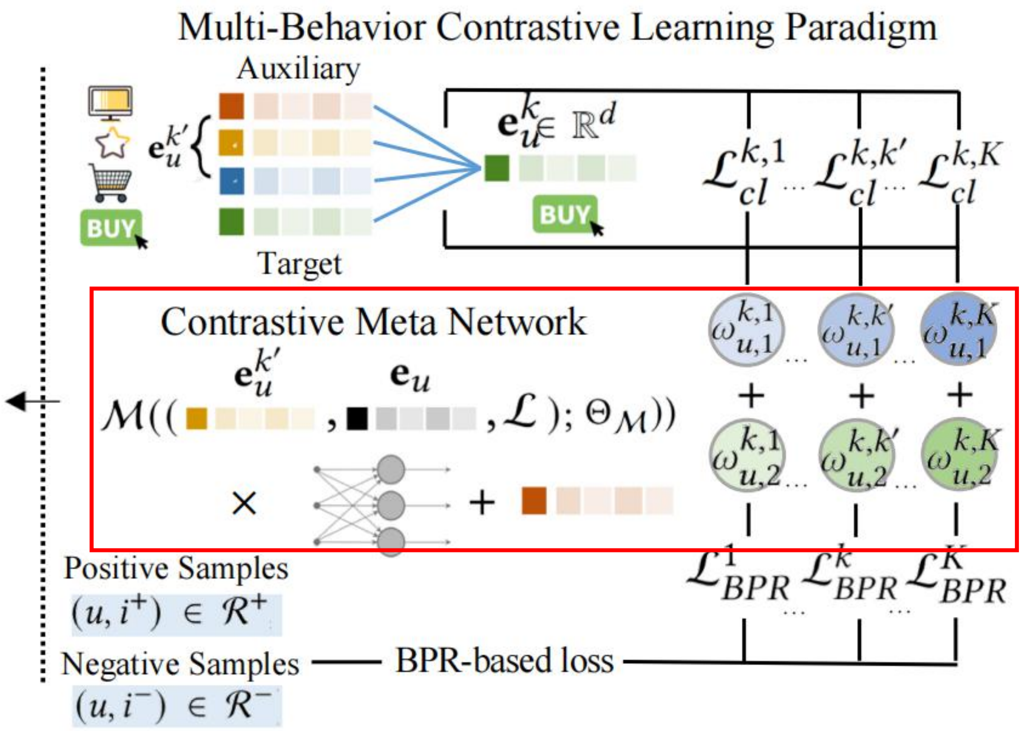
- *Meta Weight Network*

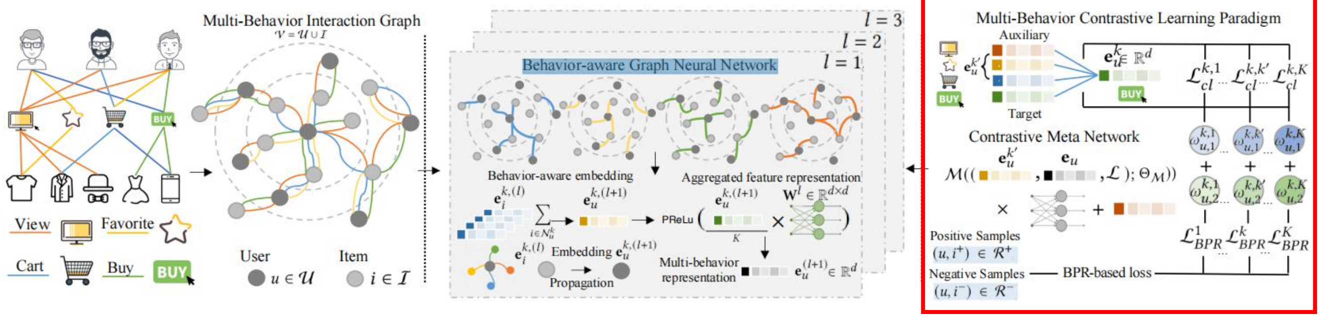
$$\xi(\mathbf{Z}_u^{k,k'}) = \text{PReLU}(\mathbf{Z}_u^{k,k'} \cdot \mathbf{W}_\xi + \mathbf{b}_\xi) \quad (5)$$

where $\mathbf{W}_\xi \in \mathbb{R}^{d \times d}$ and $\mathbf{b}_\xi \in \mathbb{R}^d$ represent the projection layer and bias term, respectively. Here, we utilize the PReLU activation function to incorporate non-linearity. On the basis of our meta weight network, we can obtain our personalized contrastive loss weight as follows:

$$\omega_u^{k,k'} = \omega_{u,1}^{k,k'} + \omega_{u,2}^{k,k'} = \xi(\mathbf{Z}_{u,1}^{k,k'}) + \xi(\mathbf{Z}_{u,2}^{k,k'}) \quad (6)$$





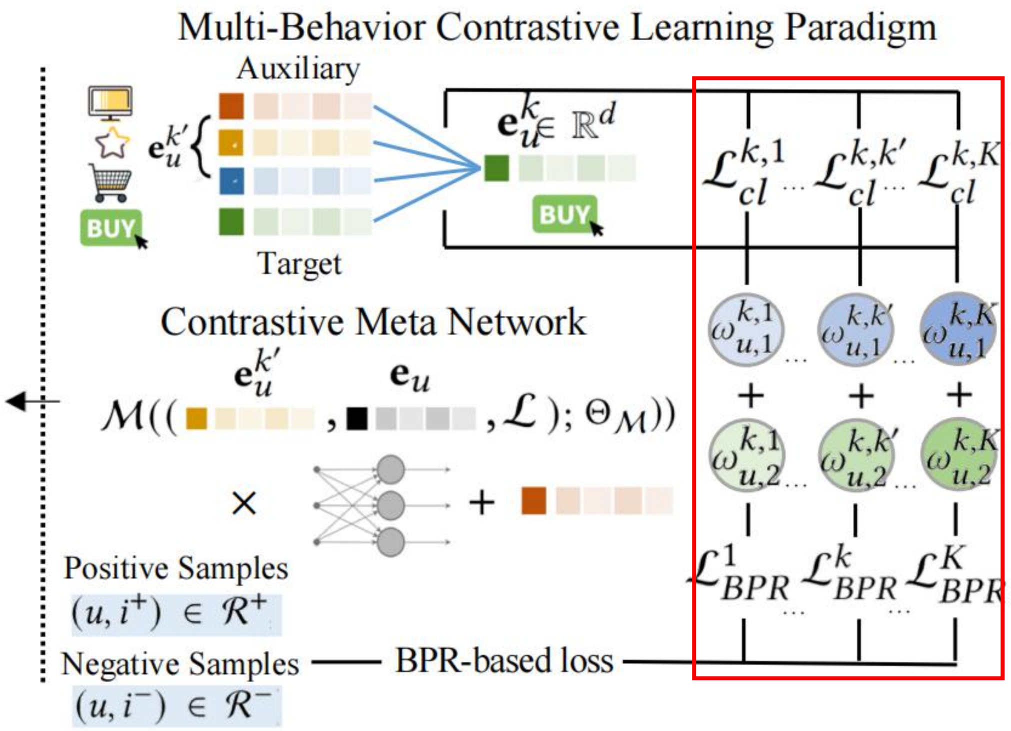


Optimization Objective

$$\mathcal{L}_{BPR}^{k} = \sum_{(u,i^+,i^-)\in O_k} -\ln(\mathrm{sigmoid}(\hat{x}_{u,i^+}^{k} - \hat{x}_{u,i^-}^{k})) + \lambda||\Theta||^2 \qquad (7)$$

O_k represents the pairwise training samples of k-th behavior type, i.e., $O_k = \{(u,i^+,i^-)|(u,i^+) \in \mathcal{R}^+, (u,i^-) \in \mathcal{R}^-\}$. Here, $\mathcal{R}^+$ and $\mathcal{R}^-$ denotes the corresponding observed and unobserved interaction of user u. Θ represents the learnable parameters and the L_2 regularization is applied for alleviating overfitting issue.

$$\mathcal{L}_{cl} = \mathcal{L}_{cl}^{k,1} + \ldots + \mathcal{L}_{cl}^{k,k'} + \ldots + \mathcal{L}_{cl}^{k,K}.$$



Experiment

Table 1: Statistics of experimented datasets

Dataset	User #	Item #	Interaction #	Interactive Behavior Type
Tmall	31,882	31,232	1,451,219	{Page View, Favorite, Cart, Purchase}
IJCAI-Contest	17,435	35,920	799,368	{Page View, Favorite, Cart, Purchase}
Retail Rocket	2,174	30,113	97,381	{Page View, Cart, Transaction}

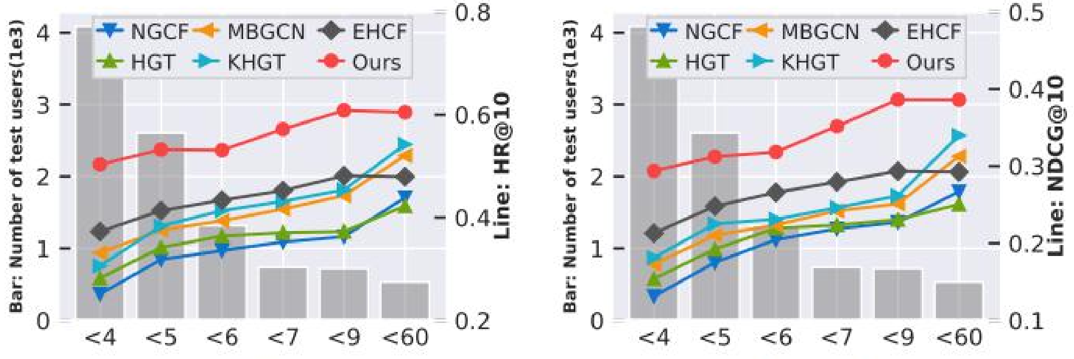
Table 2: Performance comparison of all compared methods on different datasets in terms of *NDCG@10* and *HR@10*

Dataset	Metric	BPR	PinSage	NGCF	LightGCN	SGL	HGT	HeCo	NMTR	MBGCN	MATN	KHGT	EHCF	**CML**	Imprv.	*p*-val.
Tmall	HR	0.243	0.274	0.322	0.342	0.350	0.357	0.358	0.362	0.381	0.406	0.391	<u>0.433</u>	**0.543**	25.4%	$3e^{-5}$
	NDCG	0.143	0.151	0.184	0.205	0.210	0.210	0.199	0.215	0.213	0.225	0.232	<u>0.260</u>	**0.327**	25.8%	$2e^{-4}$
IJCAI-Contest	HR	0.163	0.176	0.256	0.257	0.249	0.250	0.262	0.294	0.304	0.369	0.317	<u>0.409</u>	**0.477**	16.6%	$9e-5$
	NDCG	0.085	0.091	0.124	0.122	0.123	0.119	0.121	0.161	0.160	0.209	0.182	<u>0.237</u>	**0.283**	19.4%	$6e-3$
Retail Rocket	HR	0.235	0.247	0.260	0.261	0.263	0.305	0.297	0.314	0.308	0.301	<u>0.324</u>	0.321	**0.356**	9.9%	$1e^{-3}$
	NDCG	0.146	0.139	0.140	0.152	0.165	0.176	0.178	0.201	0.181	0.181	0.202	<u>0.207</u>	**0.222**	7.3%	$1e^{-2}$

Experiment

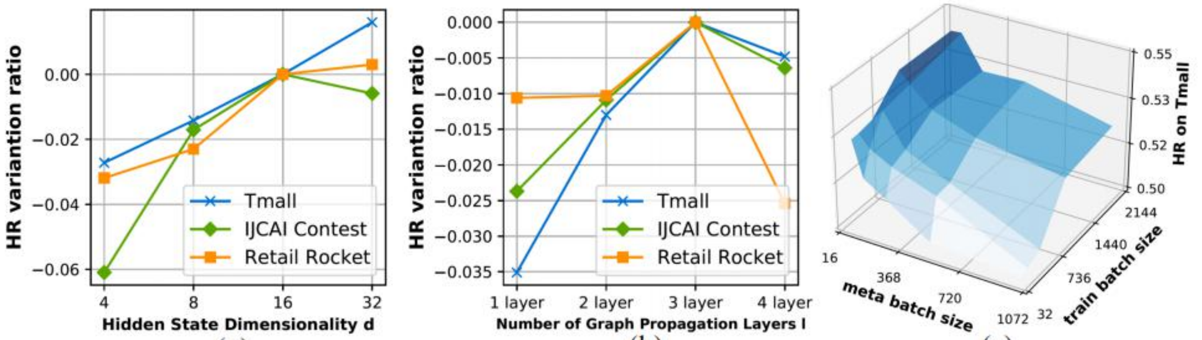
Table 3: Ablation study on key components of CML

Data	Tmall		IJCAI-Contest		Retailrocket	
Metrics	HR	NDCG	HR	NDCG	HR	NDCG
w/o-CLF	0.4665	0.2752	0.3636	0.1978	0.3032	0.1864
w/o-MCN	0.5211	0.3097	0.4527	0.2703	0.3523	0.2185
w/o-MKE	0.5237	0.2988	0.4601	0.2715	0.3506	0.2079
CML	**0.5431**	**0.3266**	**0.4769**	**0.2829**	**0.3560**	**0.2219**



(a) HR@10 (b) NDCG@10

Figure 2: Performance comparison *w.r.t* different interaction sparsity degrees on Tmall data.



(a) (b) (c)

Figure 3: Hyperparameter analysis of CML.

Experiment

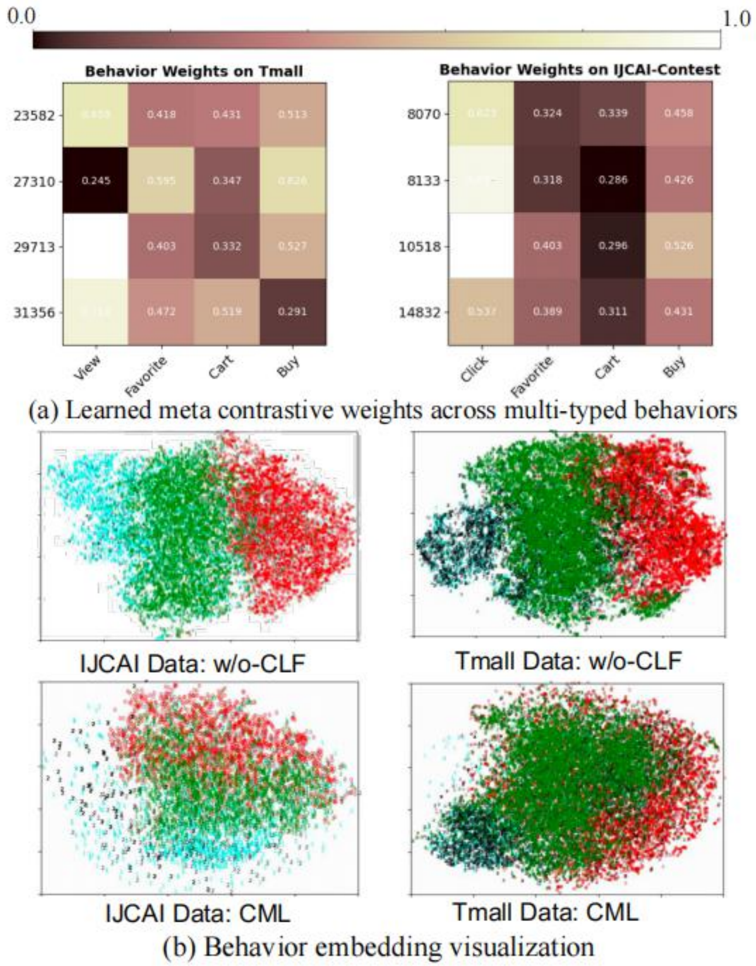


Figure 4: Model interpretation study with (a) case studies of personalized contrastive weights from sampled different users; and (b) behavior embedding visualization, *i.e.*, red: page view, blue: add-to-favorite, black: add-to-cart, green: purchase. Best viewed in colors.

Thank you!