

A Self-supervised Mixture-of-Experts Framework for Multi-behavior Recommendation



Kyungho Kim



Sunwoo Kim



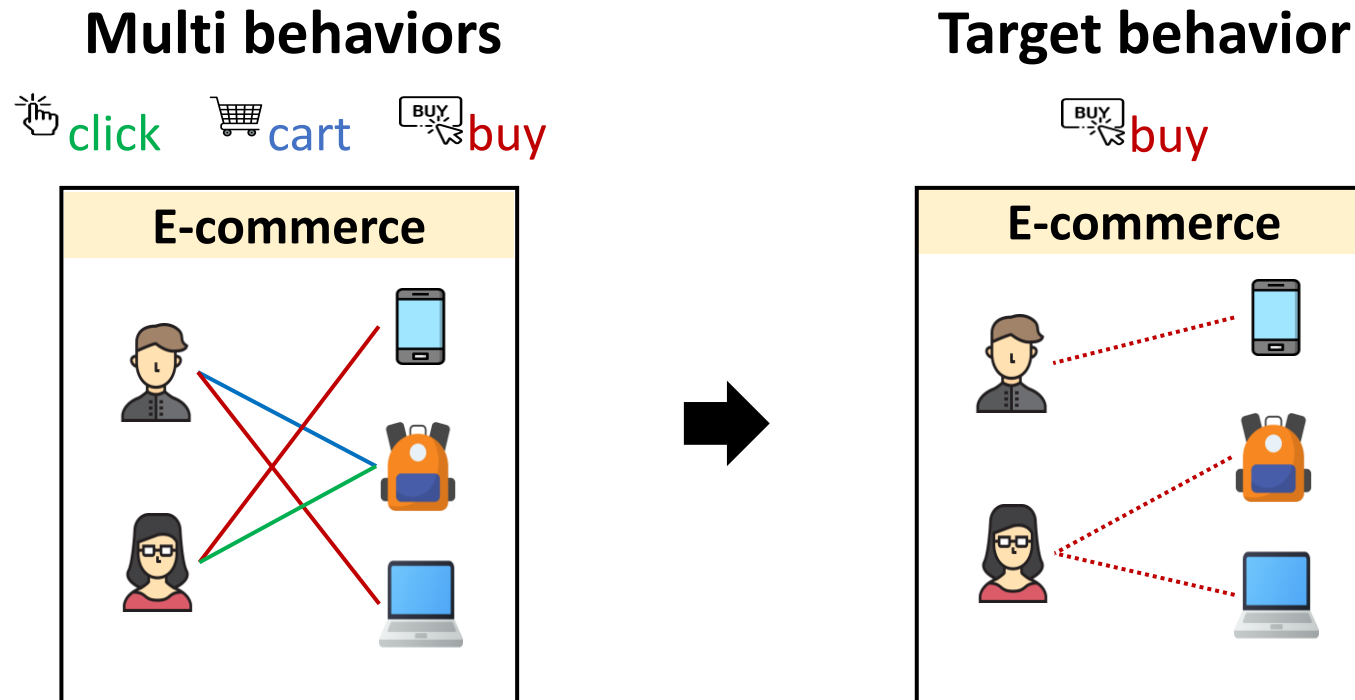
Geon Lee



Kijung Shin

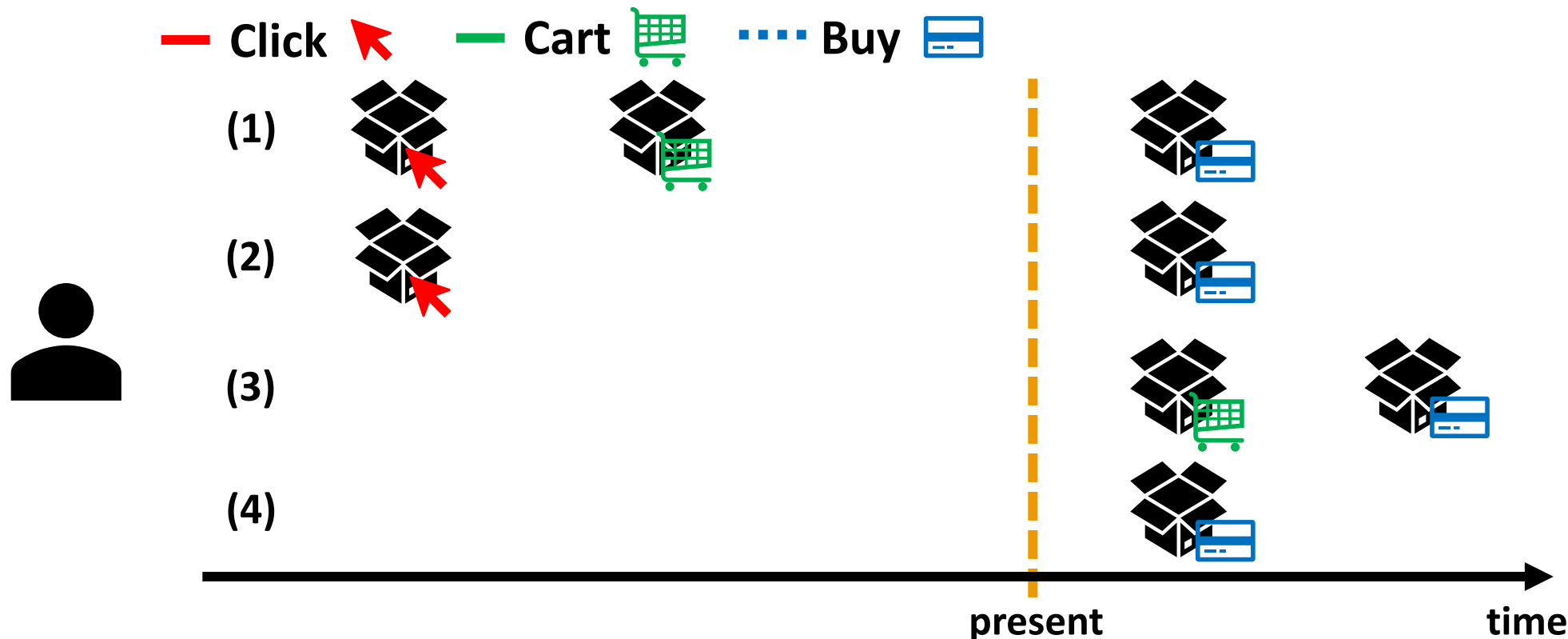
Multi-behavior Recommendation: Definition

- Users interact with items in multiple ways (e.g., by clicking or adding them to the cart).
- **Goal:** To predict future purchases by leveraging interactions of various types.



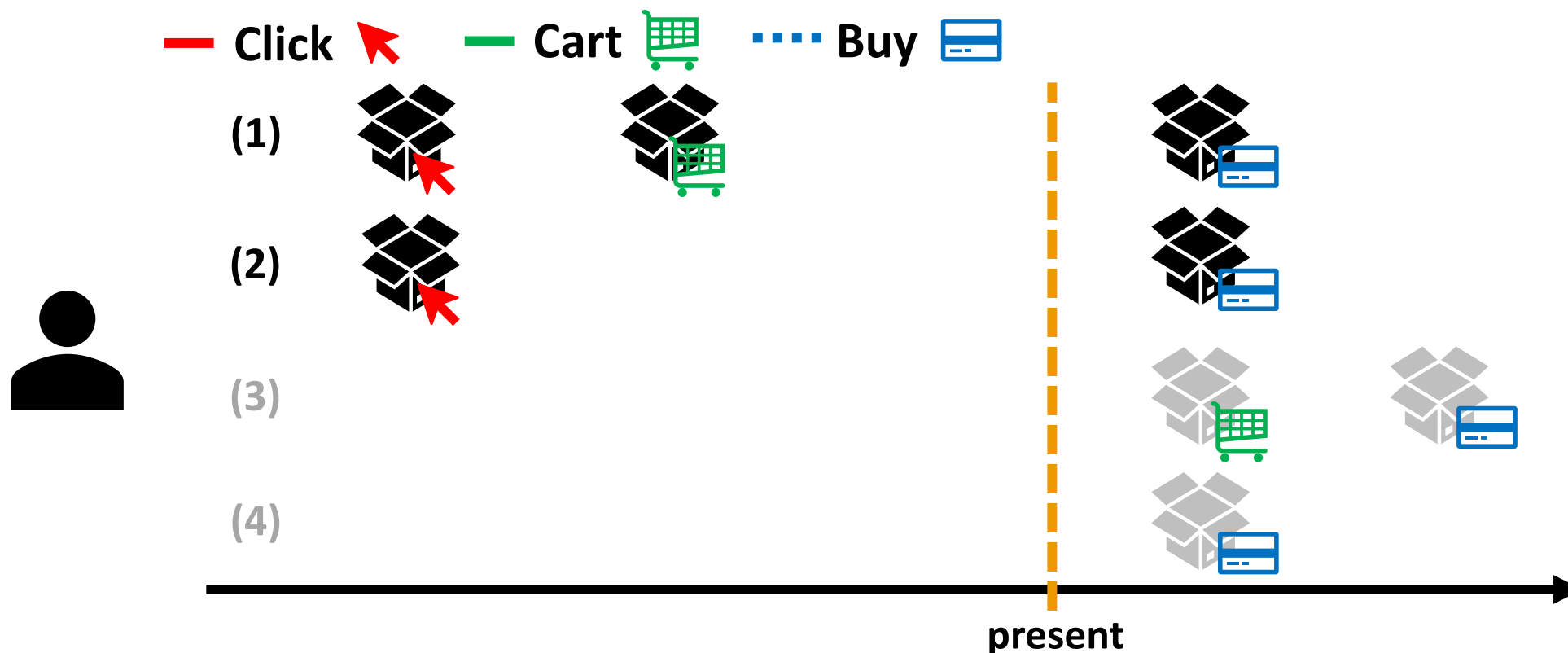
Motivational Analysis: Concepts

- For each user, we divide **unpurchased items** into:
 - Visited items:** Items that a user **has interacted with** through auxiliary behaviors.
 - Unvisited items:** Items that a user **has not interacted with** through auxiliary behaviors.



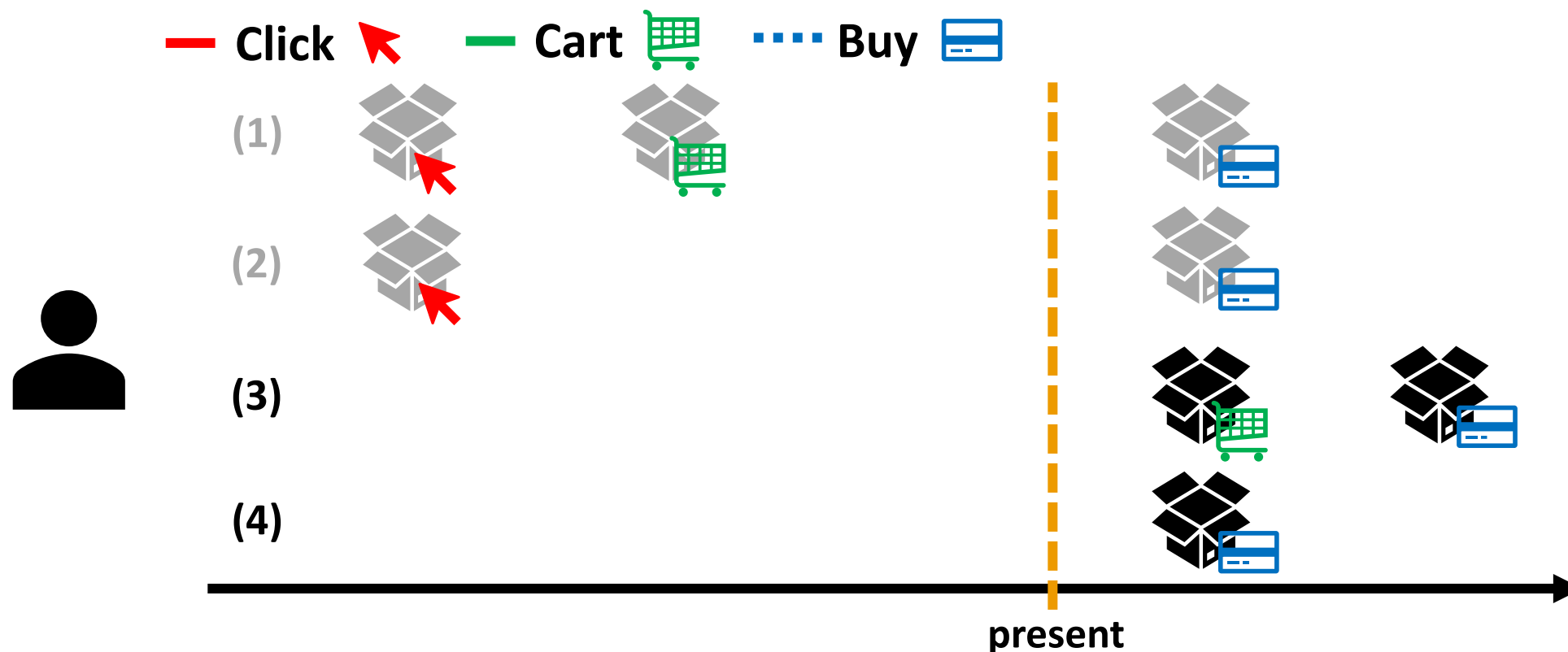
Motivational Analysis: Concepts (cont.)

- For each user, we divide **unpurchased items** into:
 - Visited items:** Items that a user **has interacted with** through auxiliary behaviors.
 - Unvisited items: Items that a user has not interacted with through auxiliary behaviors.



Motivational Analysis: Concepts (cont.)

- For each user, we divide **unpurchased items** into:
 - Visited items: Items that a user has interacted with through auxiliary behaviors.
 - Unvisited items**: Items that a user **has not interacted with** through auxiliary behaviors.

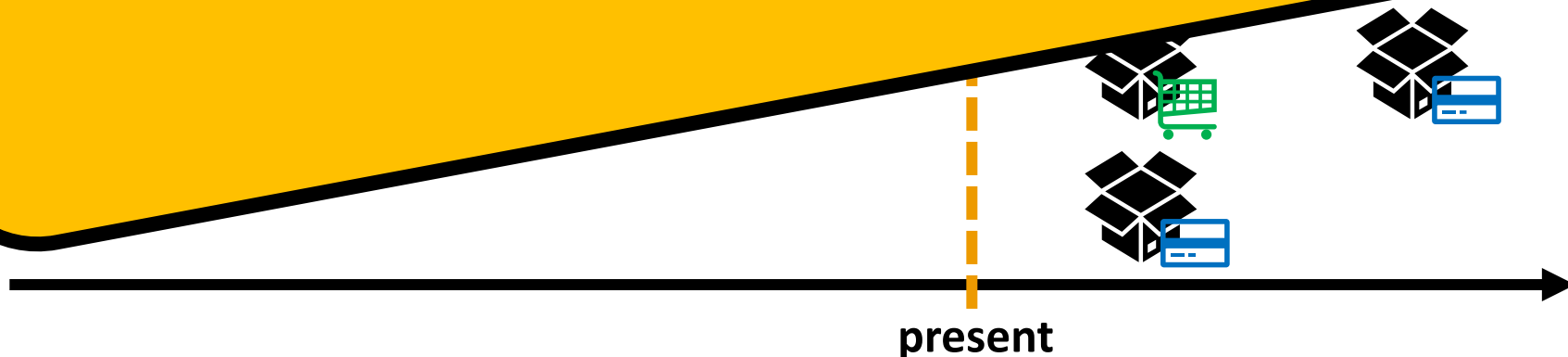


Motivational Analysis: Concepts (cont.)

- For each user, we divide **unpurchased items** into:

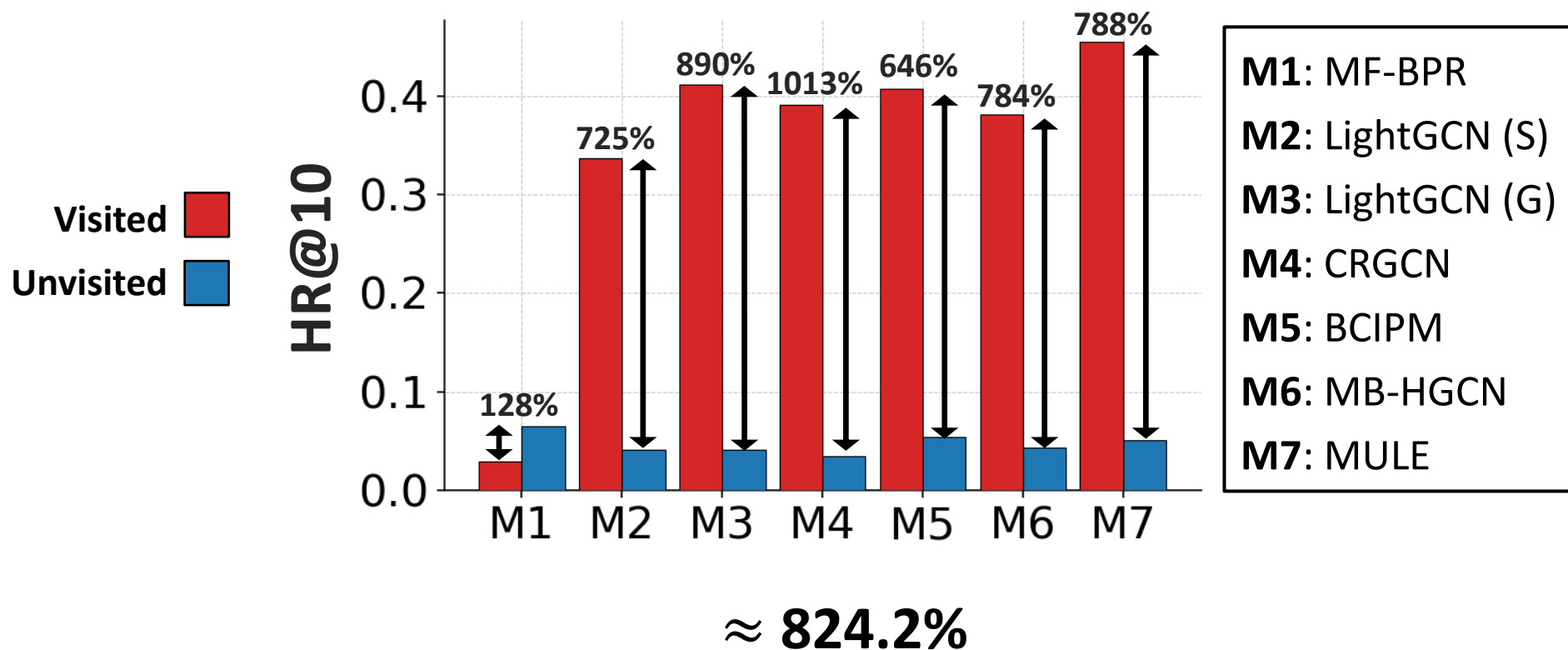
- **Visited items:** Items that a user has interacted with
- **Unvisited items:** Items that a user has not interacted with

Unvisited Items $\neq$ Cold-start Items



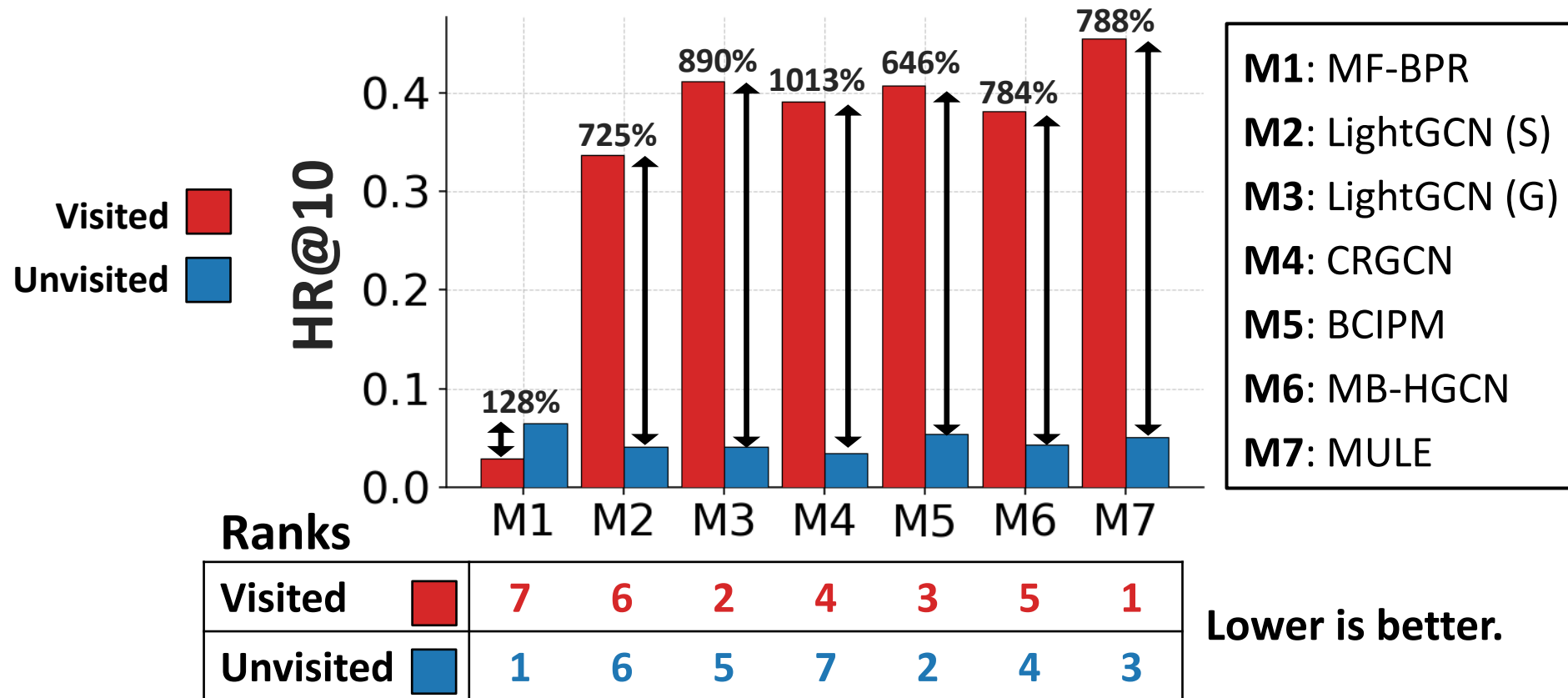
Motivational Analysis: Observations

- Our method is motivated by two observations:
 - **Observation 1.** Performance gap is large between visited and unvisited items.



Motivational Analysis: Observations (cont.)

- Our method is motivated by two observations:
 - **Observation 2.** There is no one-size-fits-all model.



Motivational Analysis: Design Rationale

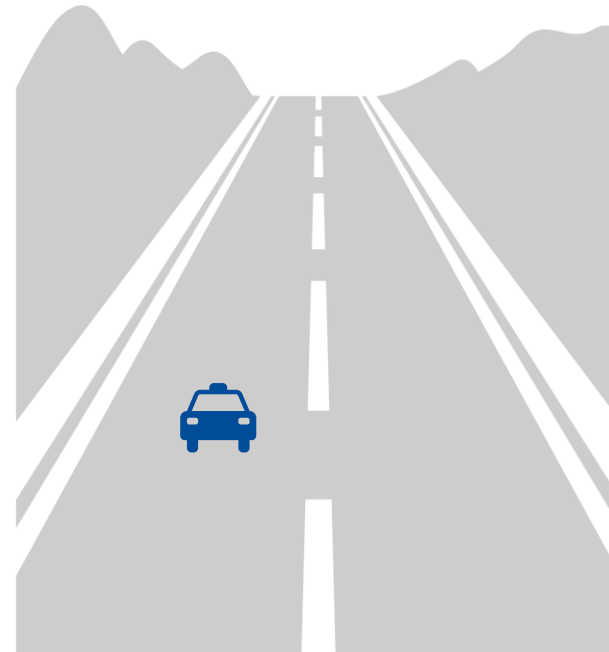
- How we addressed each observation:
 - Observation 1. Performance gap is large between visited and unvisited items.
→ **Specific focus on unvisited items.**
 - Observation 2. No one-size-fits-all model.
→ Separate experts & expert-specific training strategy.

Motivational Analysis: Design Rationale (cont.)

- How we addressed each observation:
 - Observation 1. Performance gap between visited and unvisited items.
 - Specific focus on unvisited items.
 - Observation 2. No one-size-fits-all model.
 - **Separate experts & expert-specific training strategy.**

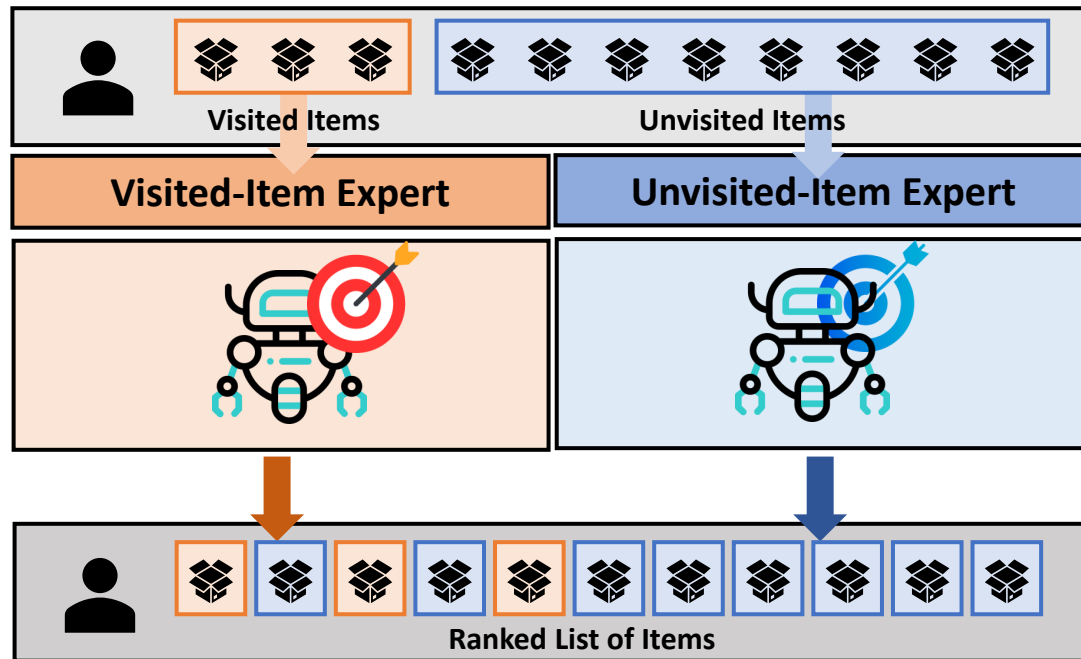
Roadmap

- **Motivational Analysis**
- **Proposed Method: MEMBER <<**
- **Experiments**
- **Conclusions**



Proposed Method: MEMBER

- MEMBER: Mixture-of-Experts for Multi-Behavior Recommendation
 - Mixture-of-Experts
 - Expertise-Adaptive Self-Supervised Learning



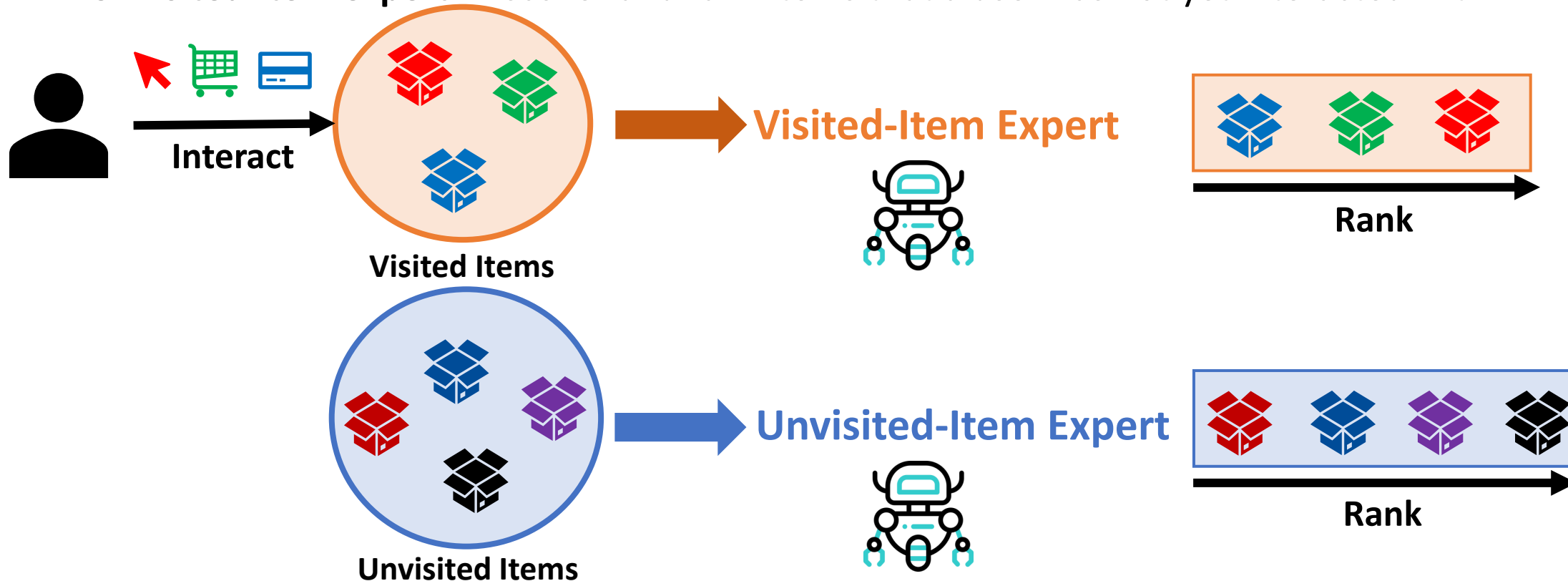
- ☒ *Visited-item recommendation*
- ☒ *Visit-filtering contrast*



- ☒ *Unvisited-item recommendation*
- ☒ *Novelty-inferring contrast*
- ☒ *Behavior-generation*

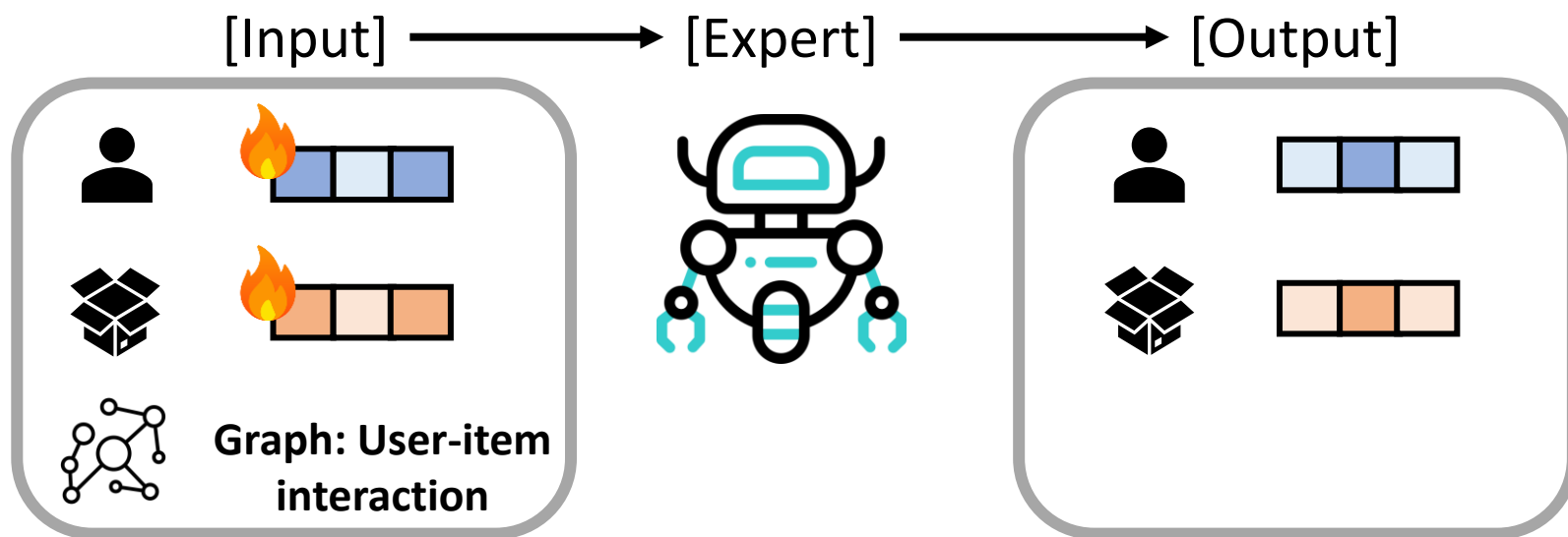
MEMBER: Mixture-of-Experts (MoE) Framework

- Mixture-of-Experts Framework
 - **Visited-item expert:** Rank items with which a user has already interacted.
 - **Unvisited-item expert:** Discover and rank items that a user has not yet interacted with.



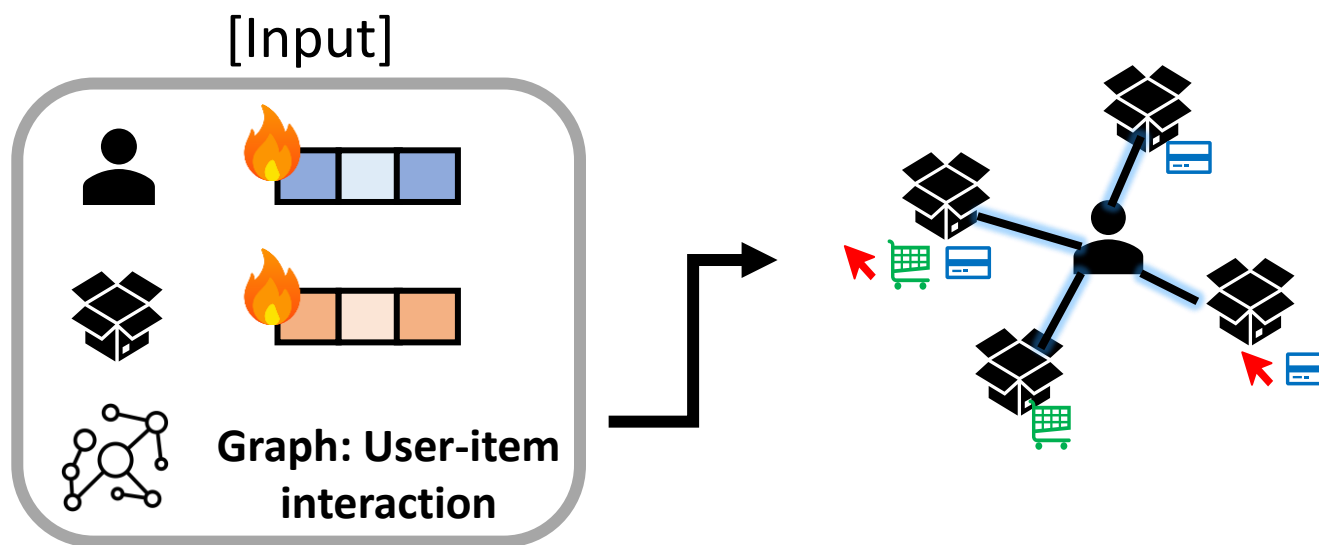
MoE Framework: Expert Design

- Each expert **encodes users and items into embeddings** based on the graph topology.
- To this end, **graph convolution** is employed.



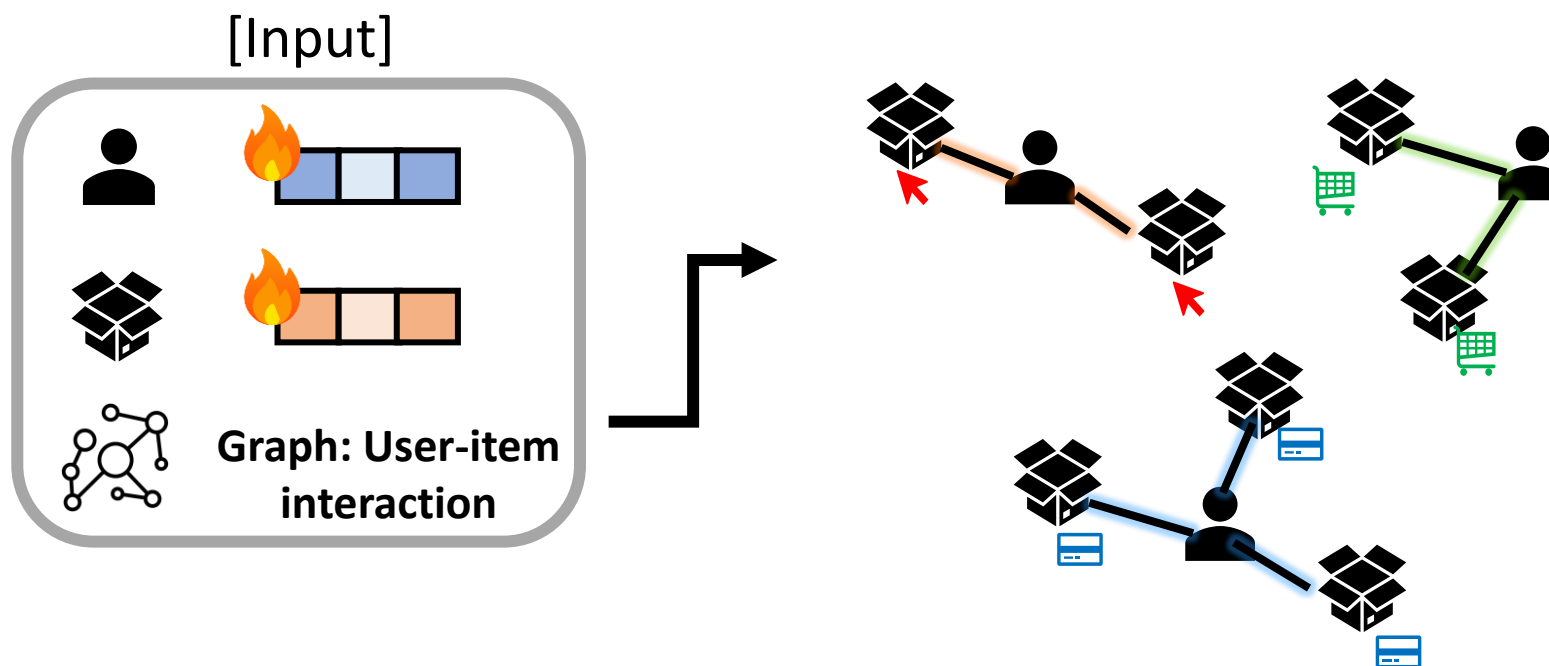
MoE Framework: Expert Design (cont.)

- Each expert has two **sub-experts** operating on distinct graphs:
 - **Global sub-expert:** Encode users and items based on **the graph including all interactions.**
 - Local sub-expert: Encode users and items based on the graph from a specific behavior.



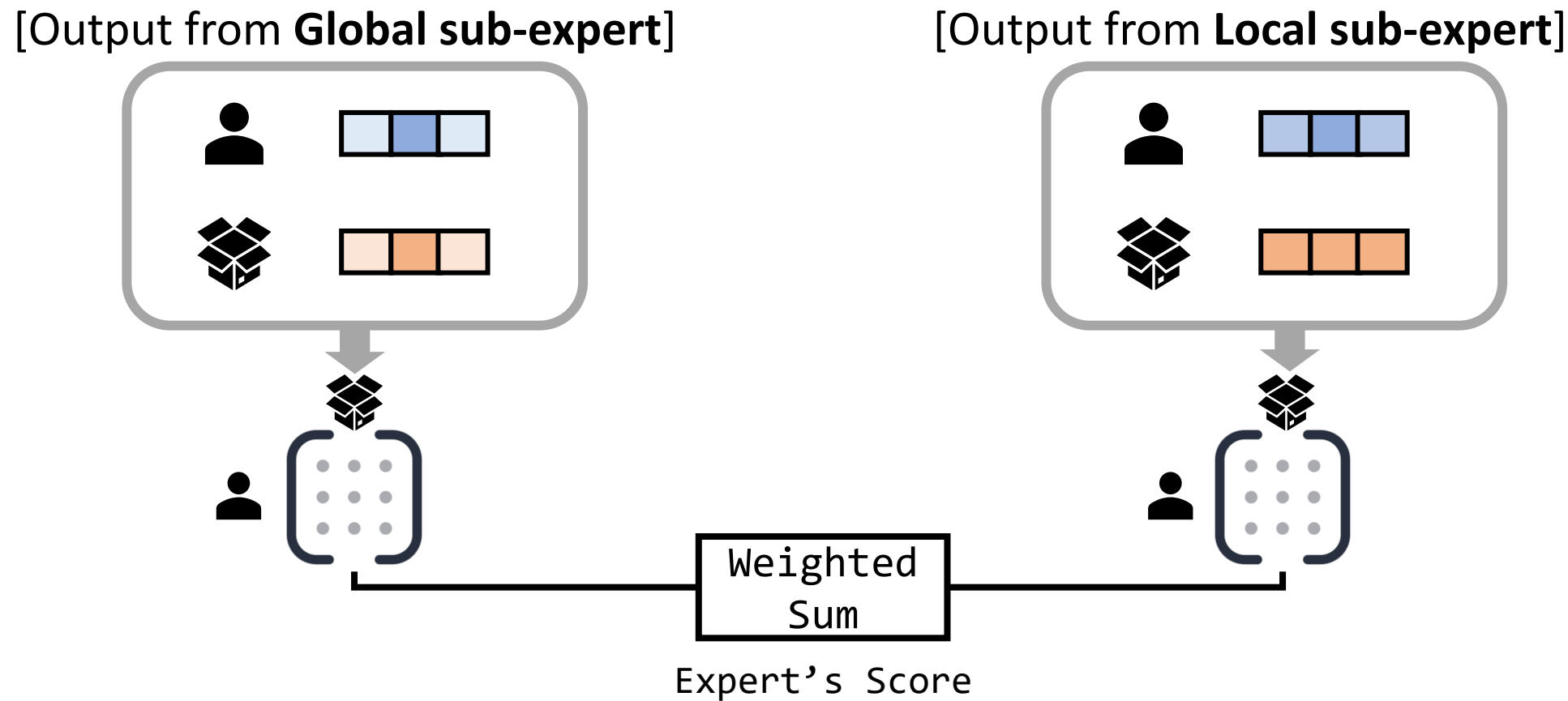
MoE Framework: Expert Design (cont.)

- Each expert has two **sub-experts** operating on distinct graphs:
 - Global sub-expert: Encode users and items based on the graph including all interactions.
 - **Local sub-expert**: Encode users and items based on the **graph from a specific behavior**.



MoE Framework: Expert Design (cont.)

- Each expert's final score is the **weighted sum of scores from two sub-experts**.



Expertise-Adaptive SSL: Overview

- Two different SSL: **(1) Contrastive learning** and **(2) Generative learning**.
 - Introduce expert-specific auxiliary losses to support the main task loss.



- ☒ *Visited-item recommendation*
- ☒ *Visit-filtering contrast <<*



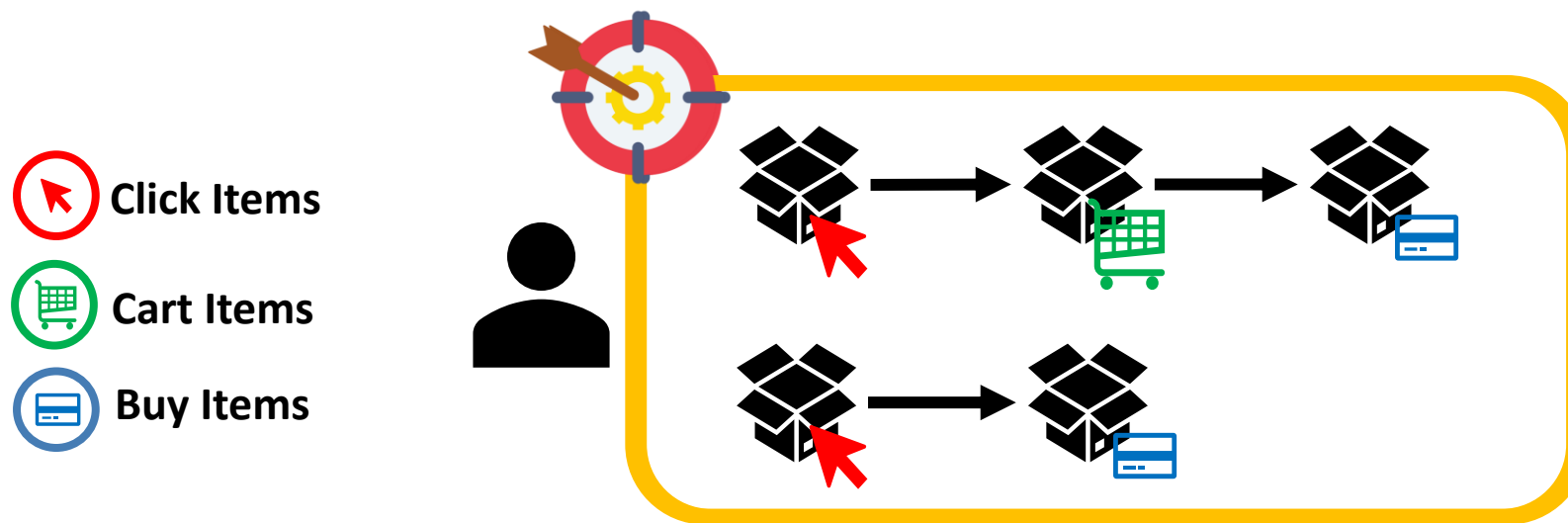
- ☒ *Unvisited-item recommendation*
- ☒ *Novelty-inferring contrast <<*
- ☒ *Behavior-generation <<*

Expertise-Adaptive SSL: (1) Contrastive Learning

- Expertise-Adaptive SSL: **(1) Contrastive Learning**
 - **Visit-filtering contrastive learning** for the visited-item expert.
 - **Novelty-inferring contrastive learning** for the unvisited-item expert.
- Goal of the expertise-adaptive contrastive learning:
 - Make expert pay enough attention to the **portion of interactions** most relevant to their task.

Contrastive Learning: Goal

- Visited-item expert: Focuses on **purchases with auxiliary interactions**.
- Unvisited-item expert: Focuses on purchases without auxiliary interactions.

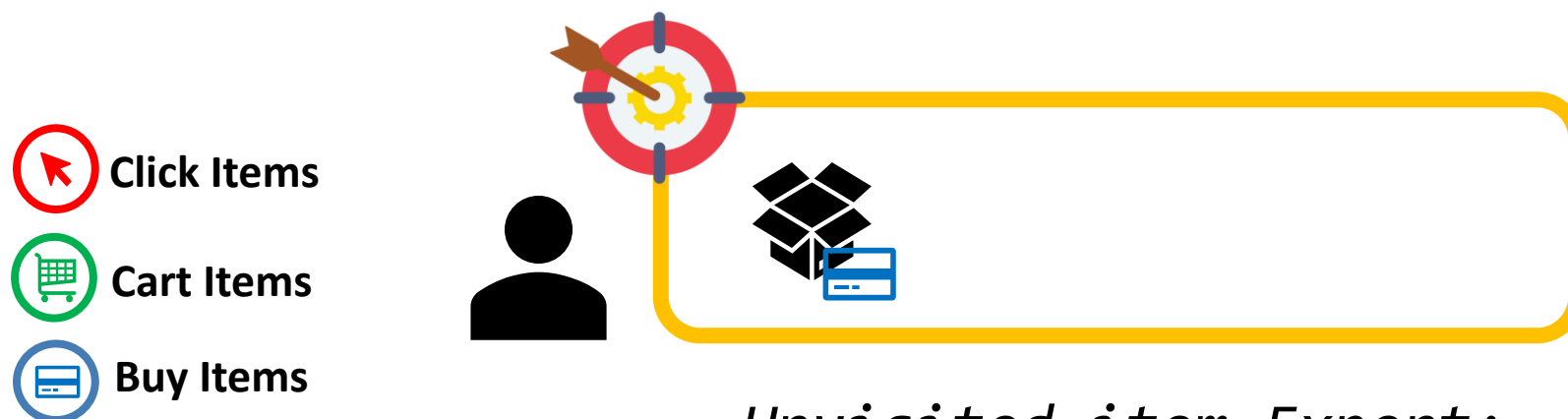


Visited-item Expert:

Should focus on these cases!

Contrastive Learning: Goal (cont.)

- Visited-item expert: Focuses on purchases with auxiliary interactions.
- Unvisited-item expert: Focuses on **purchases without auxiliary interactions.**

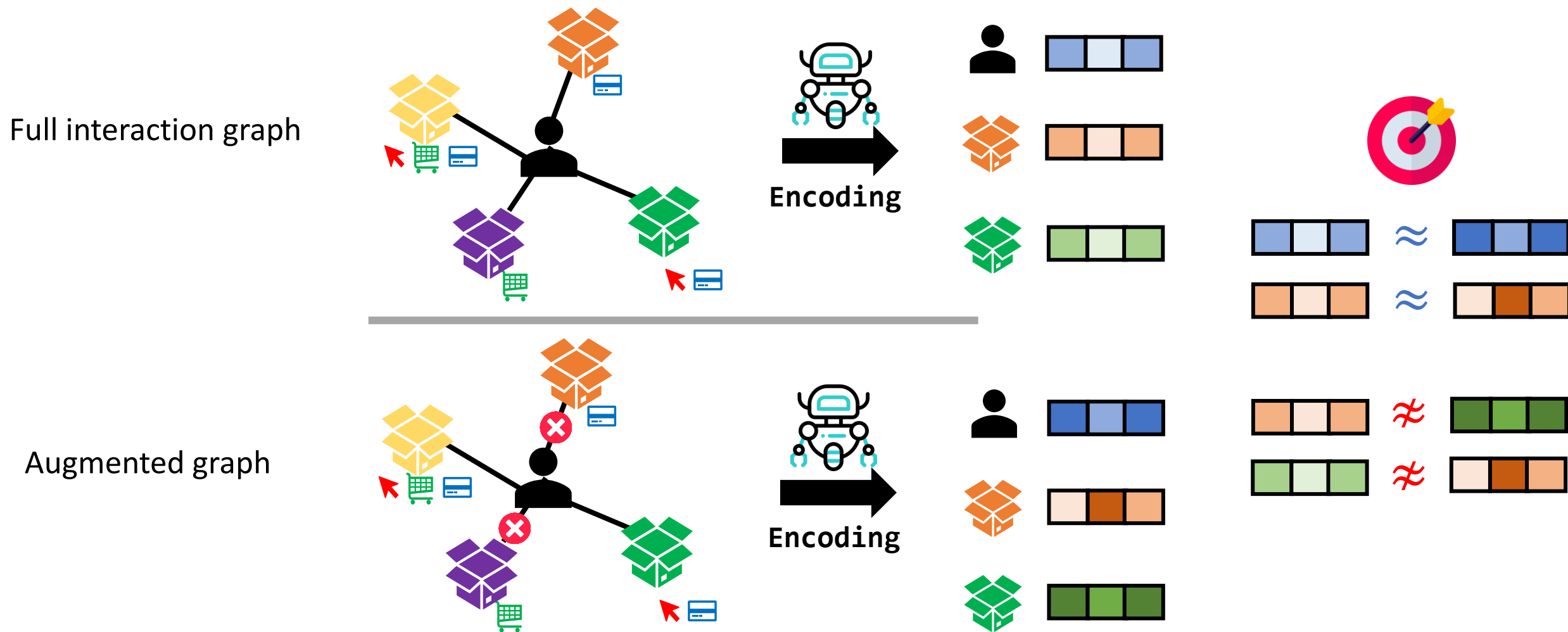


Unvisited-item Expert:

Should focus on this case!

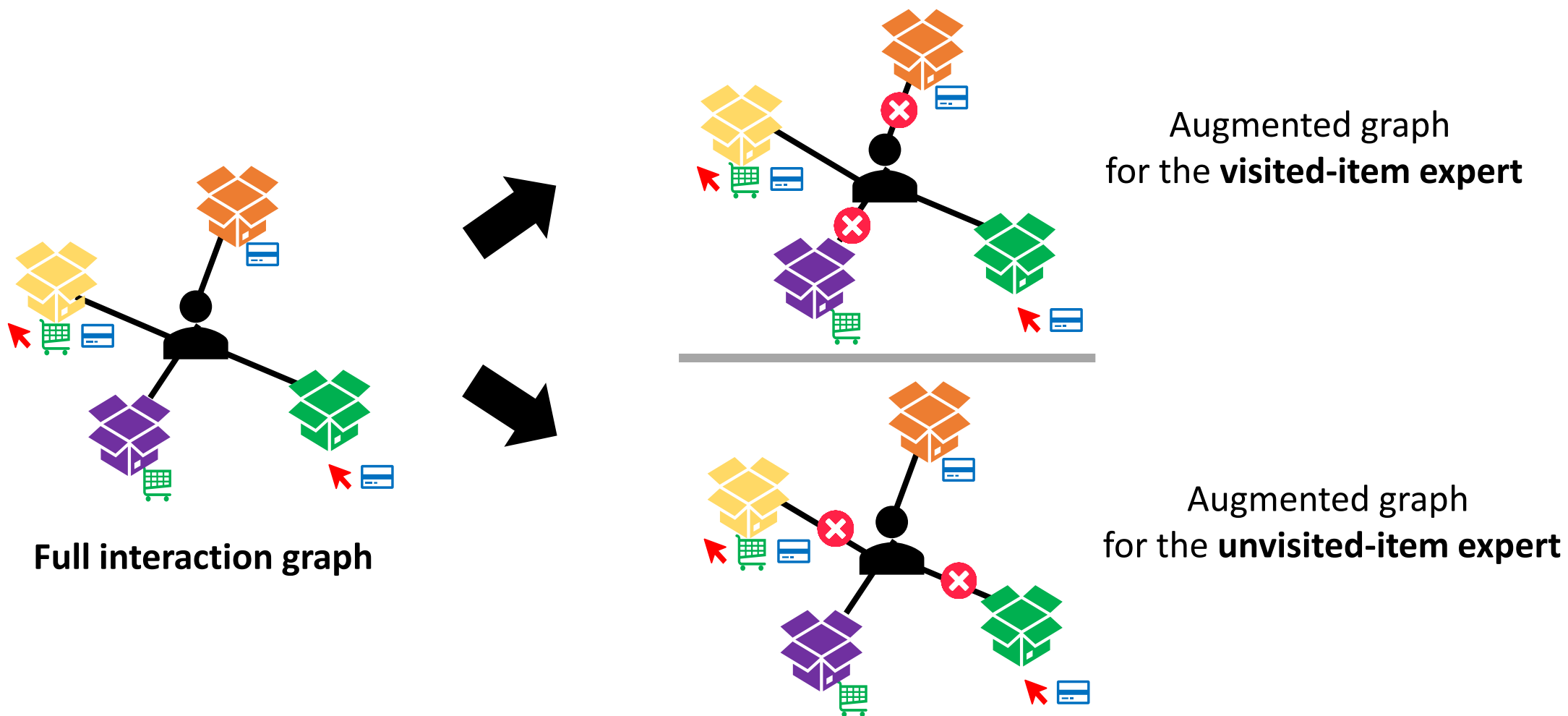
Contrastive Learning: Objective

- Contrast embeddings from a full-interaction graph and an augmented graph.



Contrastive Learning: Graph Augmentation

- For **graph augmentation**, we keep only the edges each expert should pay attention.



Expertise-Adaptive SSL: (2) Generative Learning

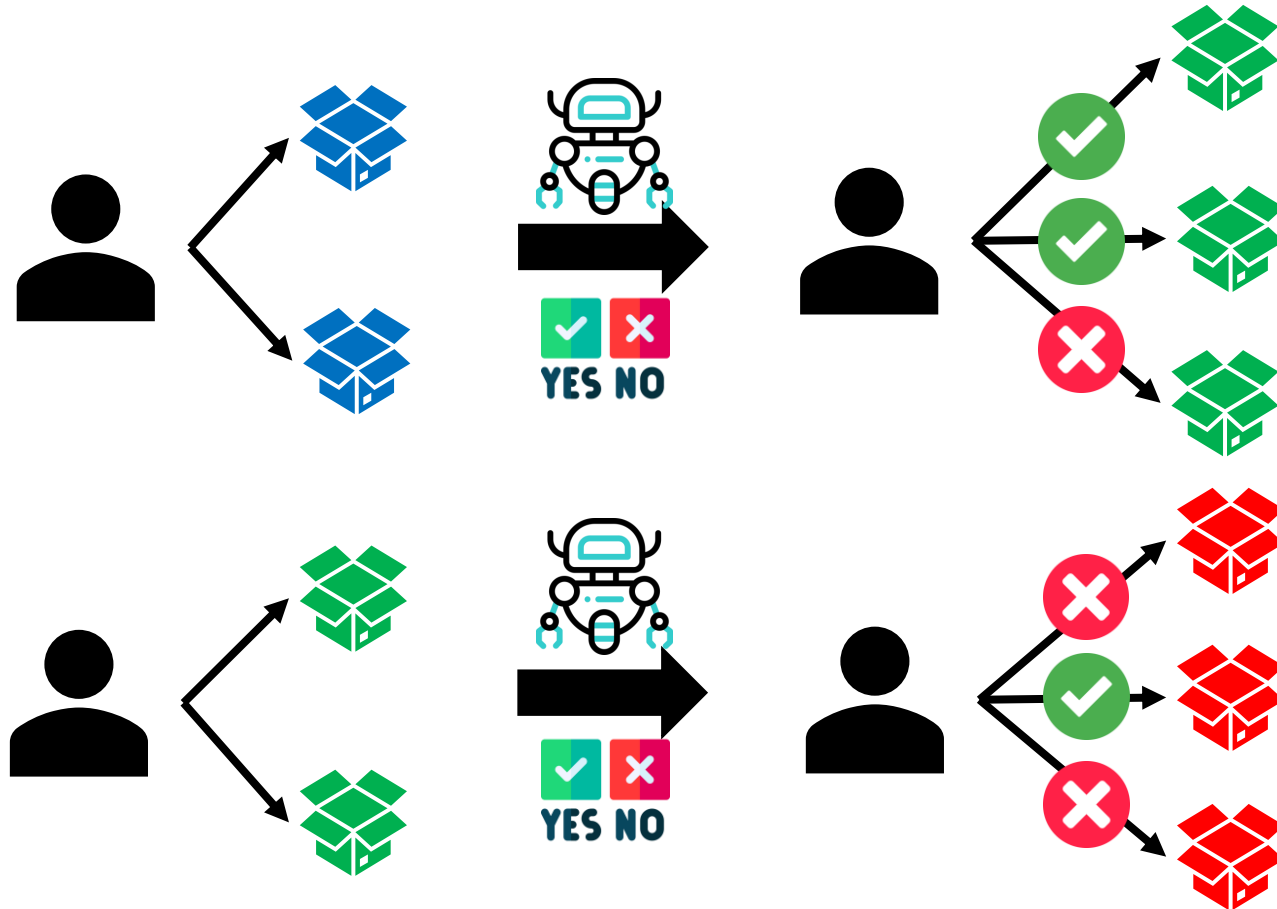
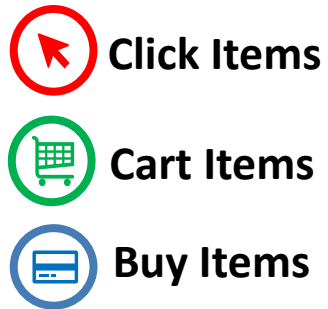
- Expertise-Adaptive SSL: (2) Generative Learning
 - **Behavior-generative learning** for the unvisited-item expert.
- Goal of the behavior-generative learning:
 - To enrich representations by predicting dense behavioral structures from sparse interactions.



Too rare!

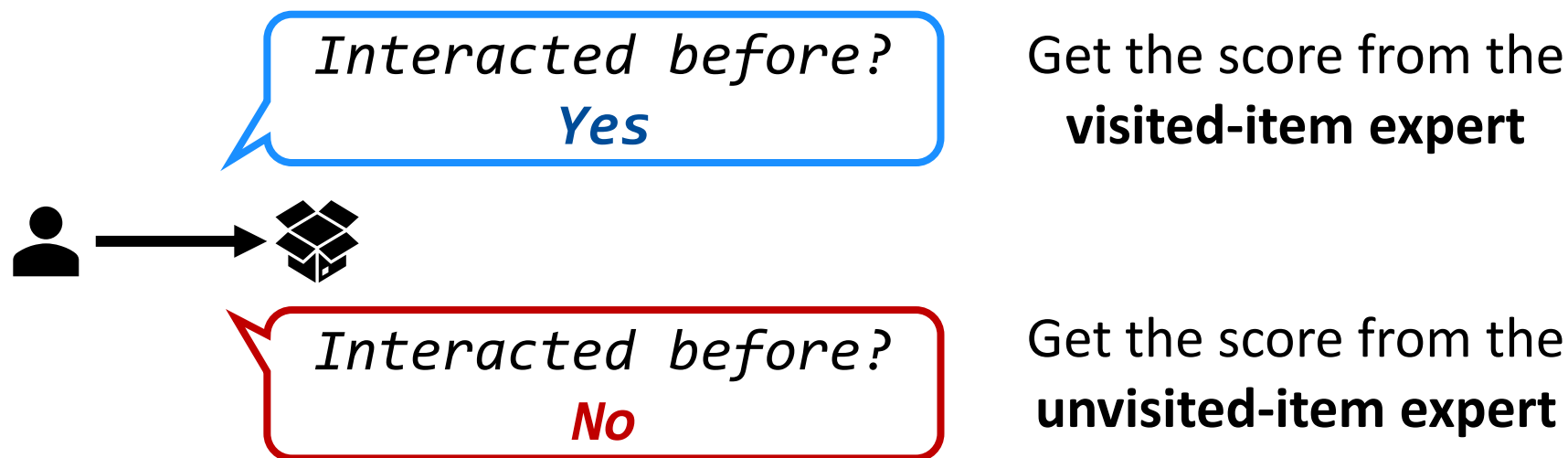
Generative Learning: Objective

- Behavior-generative learning for the unvisited-item expert:
 - Reconstruct dense (i.e., frequent) behaviors (i.e., edge) from sparse (i.e., rare) ones.



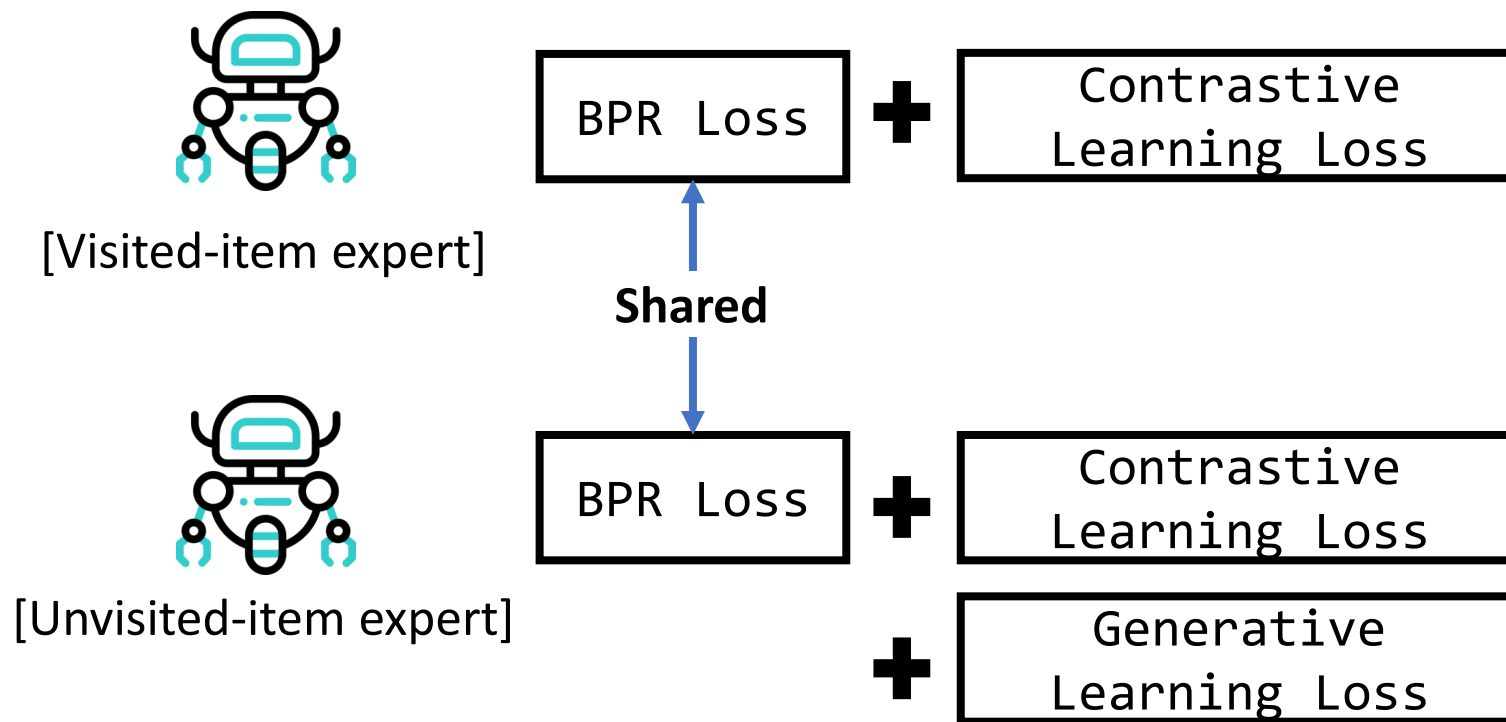
Scoring: How to Integrate Two Experts

- For each user-item pair, we choose the expert based on the **user-specific item type**.
 - **Visited** or **Unvisited**.
- The score from the chosen expert is used as the final score for the pair.



Training: How to Train Each Expert

- As training objectives, we use the weighted sum of **BPR loss** and **expertise-specific SSL losses**.
 - BPR loss: commonly-adopted recommendation loss based on pairwise comparisons.



Roadmap

- **Motivational Analysis**
- **Proposed Method: MEMBER**
- **Experiments <<**
- **Conclusions**



Experiment Settings

- **Datasets:** Three benchmark datasets on the e-commerce (Tmall, Taobao, Jdata).
- **Baselines:** Nine baseline methods.
 - Single-behavior recommender systems: MF-BPR, LightGCN
 - Multi-behavior recommender systems: MBGMN, CML, CIGF, CRGCN, BCIPM, MBHGCN, MULE
- **Evaluation protocols:** Two evaluation protocols.
 - Standard setting: Overall recommendation over all items.
 - Item-type specific settings: Separate evaluation on visited vs. unvisited items.

Experimental Results: Standard Setting

- In the **standard recommendation** setting:
 - **MEMBER** achieves **SOTA performance** across all three benchmarks.
 - **MEMBER** outperforms the best competitor by up to **65%** in terms of Hit Ratio@20.

Dataset	Metric	Single-Behavior			Multi-Behavior 							
		MF-BPR	LGCN	LGCN-G	MB-GMN	CML	CIGF	CRGCN	BCIPM	MB-HGCN	MULE	MEMBER (Ours)
Tmall	HR@10	0.0352	0.0208	0.1358	0.0480	0.0363	0.0443	0.1214	0.2136	0.1991	<u>0.2741</u>	0.3764 (0.005)
	NDCG@10	0.0181	0.0102	0.0734	0.0246	0.0181	0.0226	0.0655	0.1109	0.1061	<u>0.1466</u>	0.1850 (0.007)
	HR@20	0.0497	0.0307	0.1902	0.0663	0.0548	0.0711	0.1731	0.2983	0.2741	<u>0.3416</u>	0.5652 (0.005)
	NDCG@20	0.0216	0.0126	0.0867	0.0282	0.0212	0.0294	0.0782	0.1317	0.1245	<u>0.1629</u>	0.2326 (0.003)
Taobao	HR@10	0.0233	0.0133	0.1432	0.0628	0.0297	0.0644	0.0933	0.2377	0.1392	<u>0.2548</u>	0.3371 (0.017)
	NDCG@10	0.0126	0.0071	0.0825	0.0373	0.0149	0.0353	0.0536	0.1276	0.0767	<u>0.1402</u>	0.1808 (0.013)
	HR@20	0.0338	0.0197	0.1858	0.0528	0.0507	0.0950	0.1263	<u>0.3276</u>	0.1896	0.3060	0.4677 (0.018)
	NDCG@20	0.0152	0.0087	0.0932	0.0210	0.0205	0.0558	0.0608	0.1503	0.0894	<u>0.1639</u>	0.2115 (0.009)
Jdata	HR@10	0.3119	0.3095	0.4802	0.2970	0.2302	0.3375	0.5010	0.5532	0.5338	<u>0.6073</u>	0.6618 (0.005)
	NDCG@10	0.1885	0.1758	0.2817	0.1734	0.1276	0.2195	0.2905	0.3173	0.3238	<u>0.4300</u>	0.4425 (0.003)
	HR@20	0.3826	0.3806	0.5868	0.3859	0.3112	0.4352	0.5850	0.6693	0.6450	<u>0.6790</u>	0.7608 (0.009)
	NDCG@20	0.2077	0.1946	0.3078	0.1995	0.1496	0.2423	0.3234	0.3475	0.3533	<u>0.4370</u>	0.4687 (0.003)

Experimental Results: Item-Type Specific Settings

- For both types of Items:
 - MEMBER** outperforms all the baseline methods in 9 out of 12 settings.

Dataset	Setting	Metric	Single-Behavior		Multi-Behavior 								
			MF-BPR	LGCN	LGCN-G	MB-GMN	CML	CIGF	CRGCN	BCIPM	MB-HGCN	MULE	MEMBER (Ours)
Tmall	Visited	HR@10	0.0288	0.3368	0.4120	0.2740	0.2902	0.2675	0.3906	0.4073	0.3818	<u>0.4545</u>	0.4593 (0.003)
		NDCG@10	0.0151	0.1635	0.1988	0.1234	0.1397	0.1232	0.1852	0.1967	0.1833	<u>0.2211</u>	0.2256 (0.004)
	Unvisited	HR@10	<u>0.0645</u>	0.0408	0.0526	0.0315	0.0245	0.0269	0.0351	0.0546	0.0432	0.0512	0.1024 (0.002)
		NDCG@10	0.0215	0.0213	0.0289	0.0167	0.0131	0.0133	0.0181	<u>0.0293</u>	0.0220	0.0273	0.0661 (0.003)
Taobao	Visited	HR@10	0.0265	0.3782	0.5058	0.4136	0.4029	0.3895	0.4934	0.4935	0.4843	<u>0.5559</u>	0.5614 (0.019)
		NDCG@10	0.0142	0.1934	0.2557	0.1918	0.2014	0.1814	0.2519	0.2450	0.2464	<u>0.2956</u>	0.2985 (0.016)
	Unvisited	HR@10	0.0176	0.0097	0.0140	0.0196	0.0134	0.0139	0.0157	0.0154	0.0197	<u>0.0242</u>	0.0272 (0.001)
		NDCG@10	0.0097	0.0054	0.0079	0.0101	0.0064	0.0073	0.0082	0.0083	0.0107	<u>0.0137</u>	0.0150 (0.001)
Jdata	Visited	HR@10	0.3075	0.6935	0.7315	0.5849	0.5678	0.5748	0.7114	0.7278	0.7164	0.7605	<u>0.7574</u> (0.001)
		NDCG@10	0.1859	0.4392	0.4499	0.2795	0.2715	0.2781	0.4314	0.4340	0.4325	<u>0.4739</u>	0.5088 (0.003)
	Unvisited	HR@10	0.3225	0.3363	0.3655	0.2408	0.1687	0.2276	0.2584	0.3645	0.3910	0.4557	<u>0.4159</u> (0.009)
		NDCG@10	0.1948	0.1968	0.2202	0.1500	0.0930	0.1428	0.1393	0.2172	0.2376	0.3139	<u>0.2664</u> (0.006)

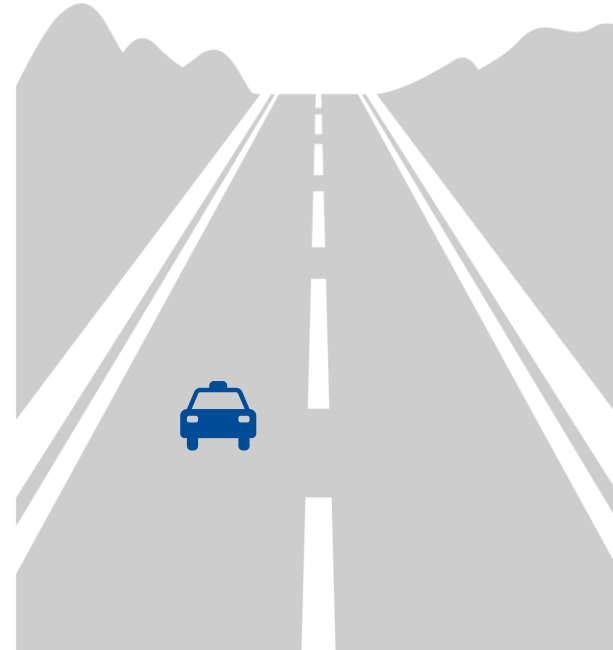
Experimental Results: Ablation Study

- Effectiveness of each component:
 - Removing **mixture-of-experts architecture** degrades performance in **all settings**.
 - Removing **SSL for visited items** degrades performance on **visited item** recommendations..
 - Removing **SSL for unvisited items** degrades performance on **unvisited item** recommendations..

Setting	Visited-item Recommendation						Unvisited-item Recommendation					
Datasets	Tmall		Taobao		JData		Tmall		Taobao		JData	
Metrics	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10
MEMBER-MoE	0.4208	0.2063	0.5516	0.2942	<u>0.7549</u>	0.5002	0.0560	0.0266	0.0218	0.0110	0.2912	0.1832
MEMBER- $\mathcal{L}_{SSL}^{(V)}$	0.4377	0.2109	0.5185	0.2692	0.7278	0.4735	<u>0.0896</u>	0.0520	<u>0.0271</u>	<u>0.0148</u>	0.4011	0.2524
MEMBER- $\mathcal{L}_{CL}^{(U)}$	0.4423	0.2167	0.5463	0.2898	0.7480	<u>0.5004</u>	0.0765	0.0459	0.0181	0.0101	<u>0.4117</u>	0.2665
MEMBER- $\mathcal{L}_{GEN}^{(U)}$	<u>0.4540</u>	<u>0.2229</u>	<u>0.5538</u>	<u>0.2975</u>	0.7428	0.4975	0.0854	<u>0.0554</u>	0.0259	0.0140	0.3883	0.2416
MEMBER	0.4593	0.2256	0.5614	0.2985	0.7574	0.5088	0.1024	0.0661	0.0272	0.0150	0.4159	<u>0.2664</u>

Roadmap

- **Motivational Analysis**
- **Proposed Method: MEMBER**
- **Experiments**
- **Conclusions <<**



Conclusion

○ Key Observations:

- (1) Performance gap is large between visited items and unvisited items.
- (2) Challenge of achieving balanced performance with a single model.

○ Novel Architecture:

- (1) Mixture-of-experts framework.
- (2) Expertise-adaptive self-supervised Learning.

○ Accuracy:

- Achieve the best performance in both standard and item-type specific settings.

Thank you!

<https://github.com/K-Kyungho/MEMBER>



Kyungho Kim



Sunwoo Kim



Geon Lee



Kijung Shin