CIPHER:
Causal Intent Plug-in Framework
for the Mitigation of Historical Exposure Bias in Recommender Systems

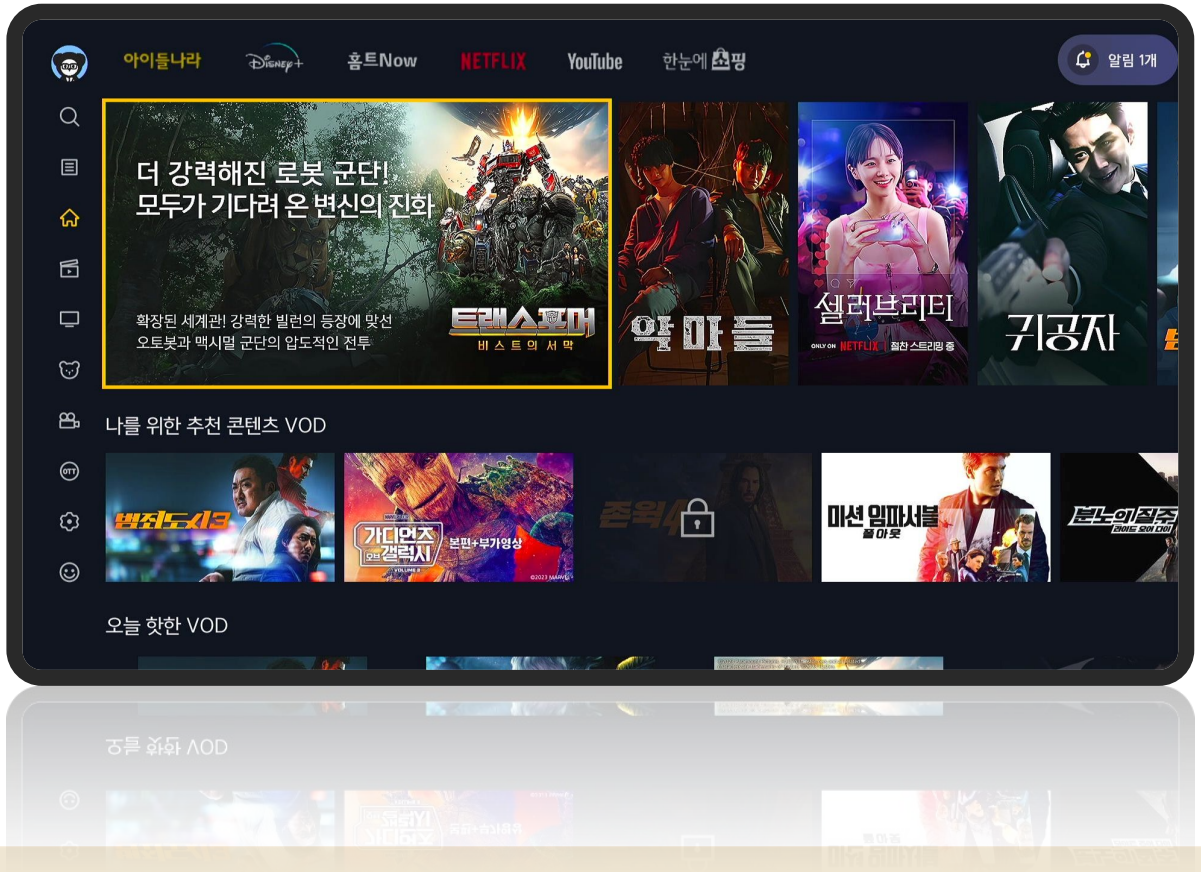
Sanghyeon Lee, Yeonghwan Jeon, Hyuncheol Jo, and Byoung-Ki Jeon (LG Uplus Corp.)

August 4, 2025

Introduction

Background

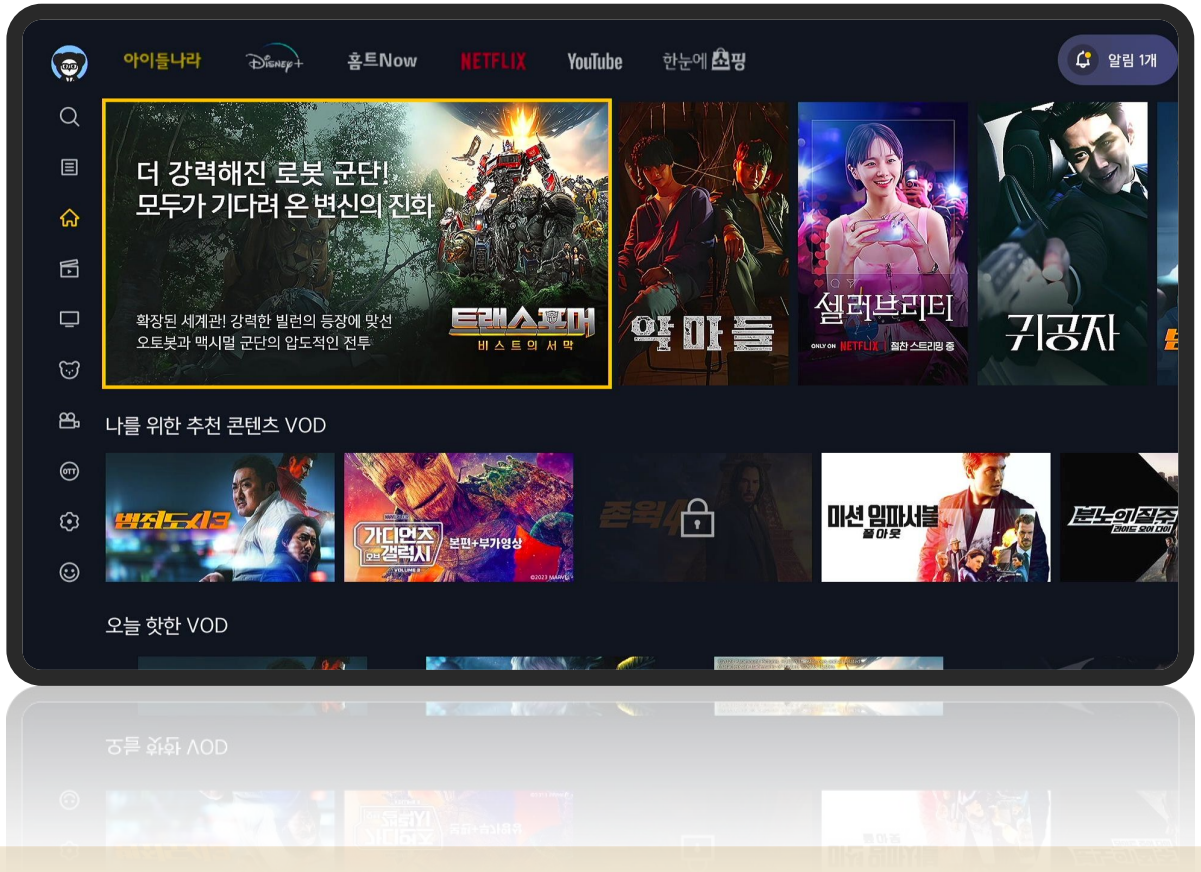
- Modern recommendation systems aim to deliver personalized content based on user interests.
- However, in practice, business priorities often lead to prioritizing popular, profitable, or newly released content.
- This results in Structural Exposure Bias within user interaction history.



Introduction

Objectives

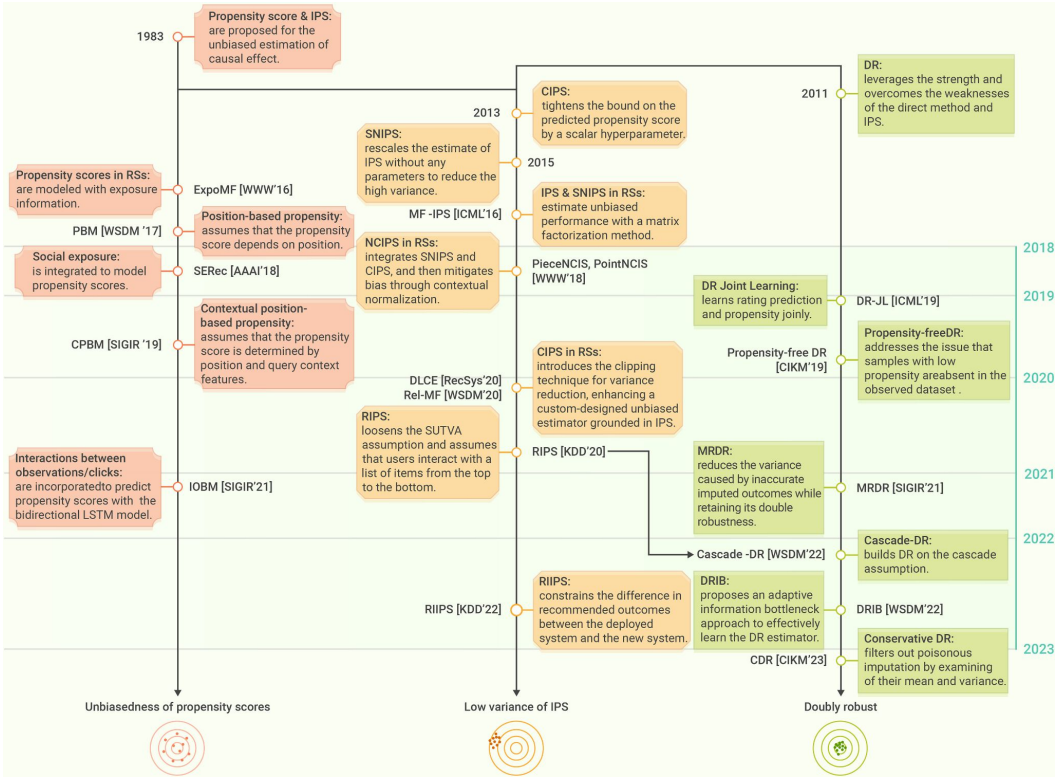
- We aimed to address Structural Exposure Bias in recommendation systems by leveraging Causal Inference.
- We sought to develop a practical, easily deployable plug-in framework to mitigate bias without altering existing recommendation system architectures.



Related Work

Existing Approaches

- Previous methods typically use propensity scores, Inverse Propensity Scoring (IPS), or Doubly Robust estimation to adjust or create unbiased data values.
- Frameworks such as DICE and COCO-SBRS manipulate and partition data or item sequences to isolate pure user interests or item sequences devoid of individual user biases.



Related Work

Limitations

- Existing methods usually require structural changes to the core recommendation models.
- Implementing these structural changes in real-world systems introduces significant risks and high costs.

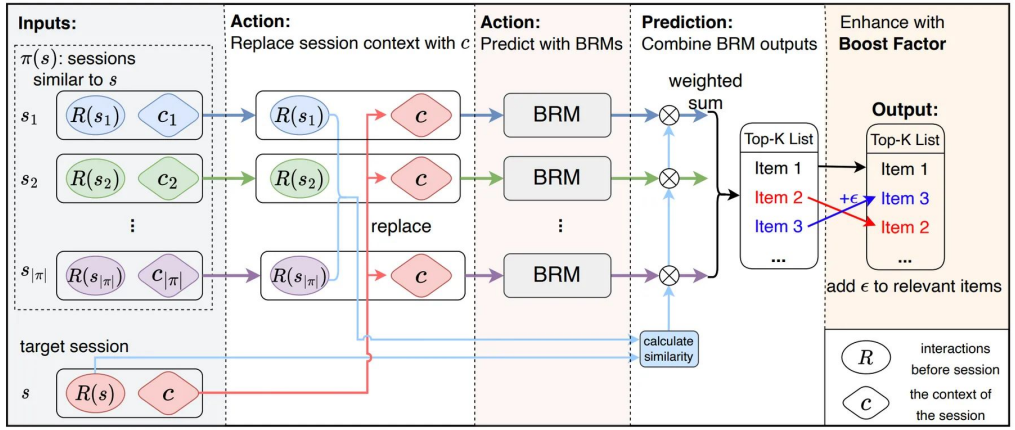
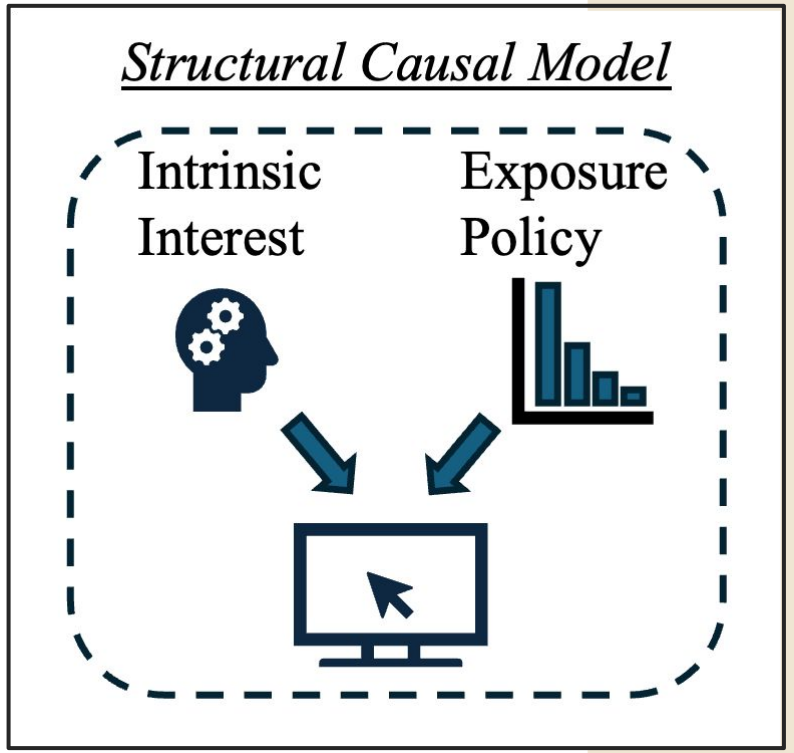


Figure 2: The steps of Action and Prediction of COCO-SBRS. The BRMs are pre-trained with the sessions in the training set in the Abduction step, which is not shown in this figure.

Proposed Method

Problem Formulation via Structural Causal Modeling

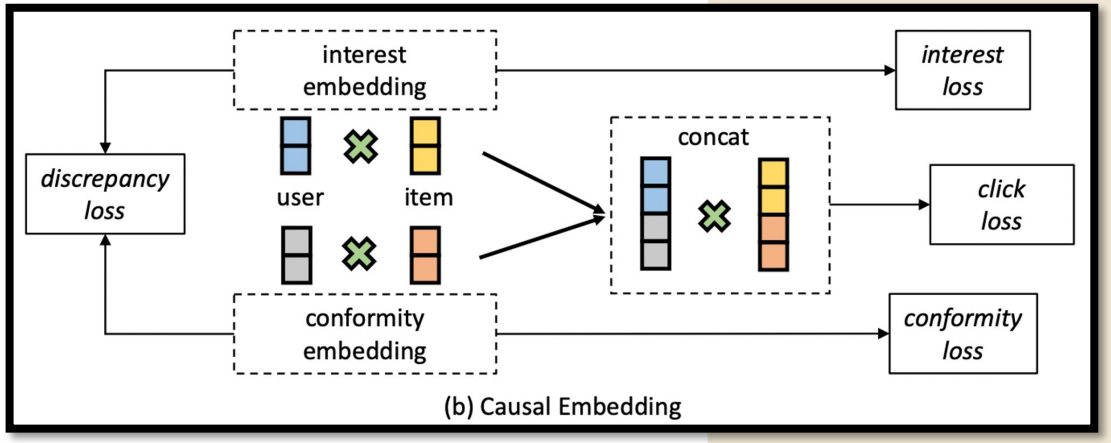
- We formulated viewing behavior (Y) as influenced by two causal factors: Interest and Conformity, creating a collider structure for causal inference.
- Datasets were partitioned to independently learn interest-driven and conformity-driven aspects.



Proposed Method

Causal Intent Learner via Disentangled Representation

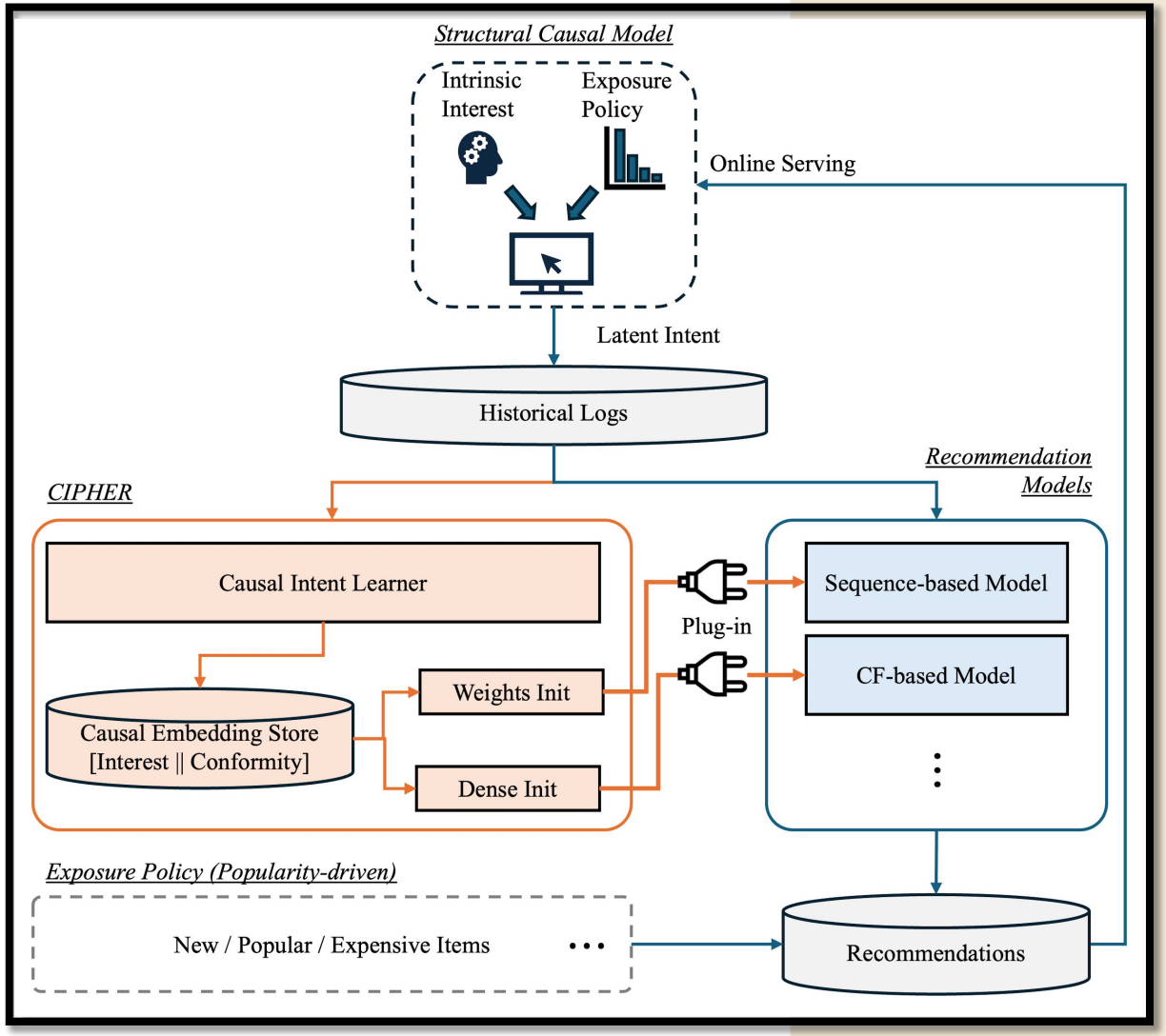
- The DICE framework was utilized to separately learn intermediate user/item embeddings representing interest and conformity.
- Multi-task learning was employed across the partitioned datasets to simultaneously optimize these embeddings.



Proposed Method

Plug-in Adaptation for Recommendation Models

- Learned causal embeddings were applied as initial weights for users and items, tailoring integration methods specifically to different model architectures.
- The model-agnostic integration allowed rapid incorporation of genuine user interests and conformity across various recommendation models.



Experiments

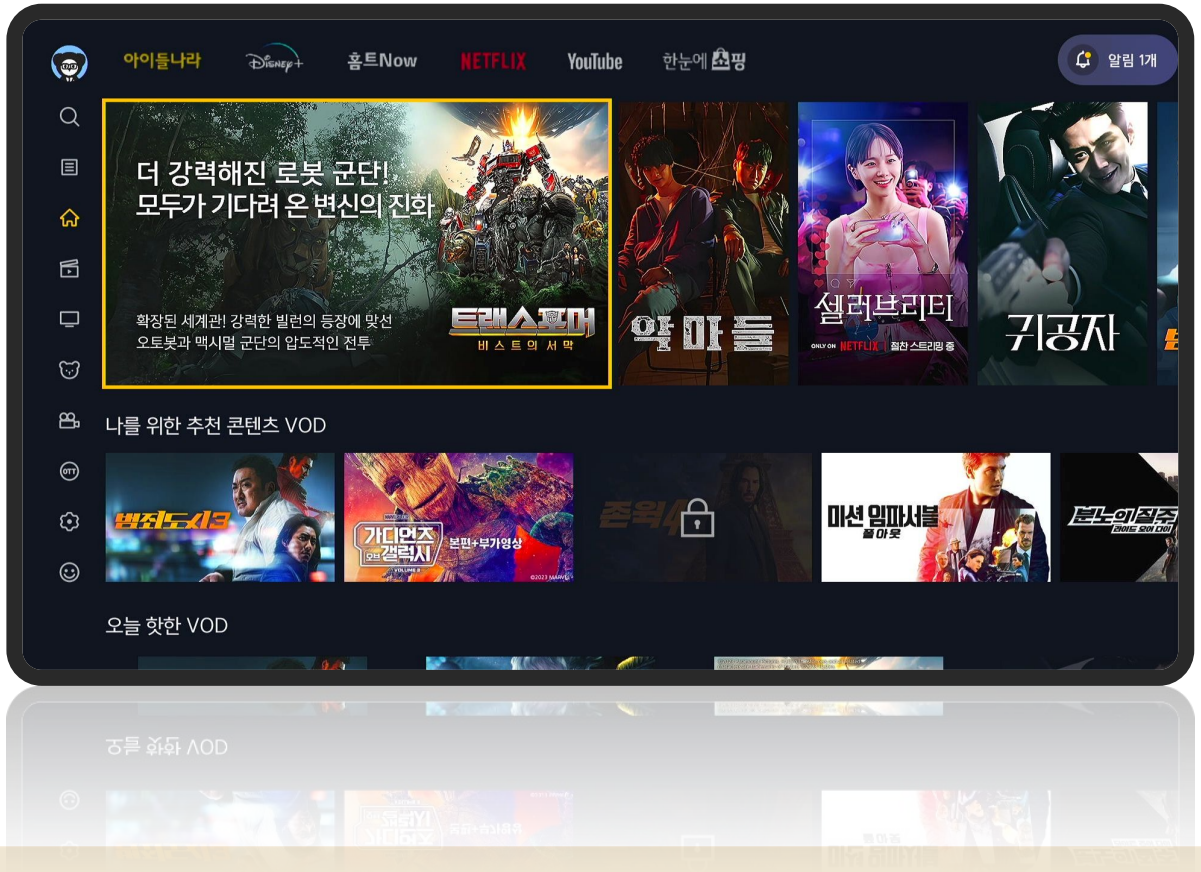
Research Objectives

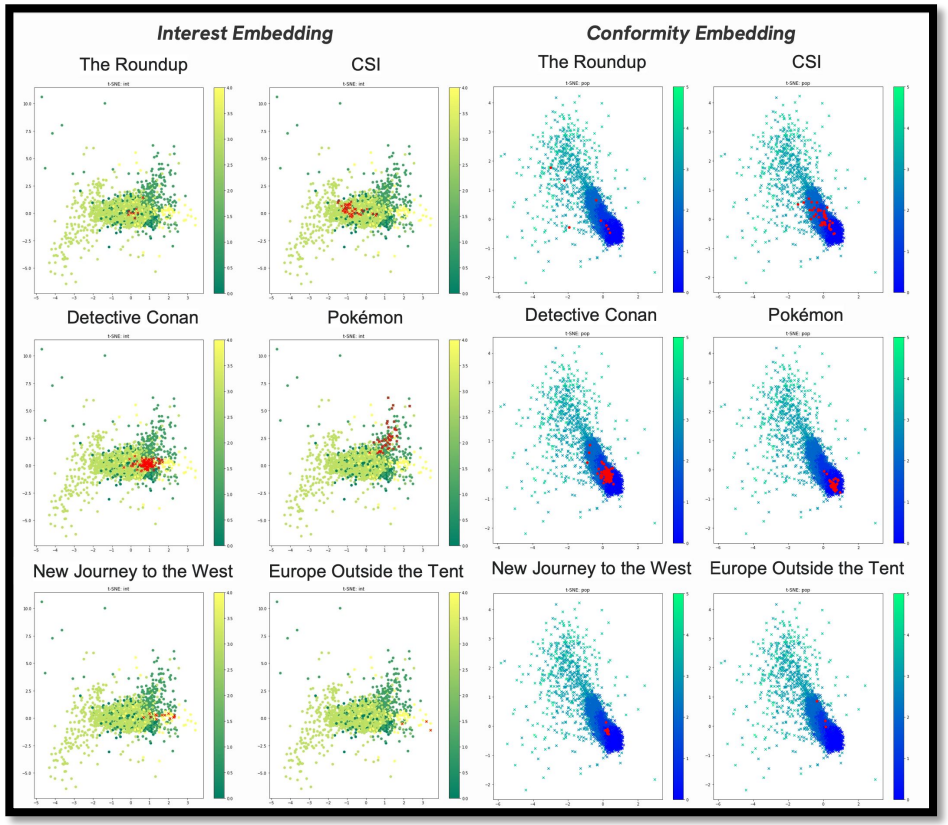
- RQ1: Can we causally disentangle user intent into intrinsic interest and conformity, producing interpretable representations from real-world interaction logs?
- RQ2: Does our integration of CIPHER embeddings into various backbone models improve recommendation performance without requiring architectural modifications?
- RQ3: How effectively does CIPHER mitigate exposure bias and improve item coverage and personalization fairness?

Experiments

Evaluation Setup

- Dataset: Real-world IPTV service logs
- Metrics: Precision, Recall, MRR, Item Coverage





Experiments

RQ1: Interpretability of Causal Intent Representations with Real-World Data

- Item clustering based on viewing history demonstrated clear interest embedding groups (e.g., series, genres) visually.
- CIPHER generates causally interpretable embeddings that clearly separate user interest from conformity, validating our core assumptions with real-world data.

Model	Precision@25	Recall@25	MRR@25	Coverage@25
Sequence-based model (base)	-	-	-	-
Sequence-based model + CIPHER (fixed)	-8.79%	-7.60%	-9.60%	-21.02%
Sequence-based model + CIPHER (trainable)	+5.86%	+5.99%	+7.09%	+55.21%
CF-based model (base)	-	-	-	-
CF-based model + CIPHER (fixed)	-1.77%	-1.93%	-3.25%	-75.95%
CF-based model + CIPHER (trainable)	+4.42%	+4.02%	+4.67%	-45.43%

Experiments

RQ2: Effectiveness of Model-Agnostic Integration

- Experiments were conducted with two distinct recommendation algorithms.
- An ablation study was performed, comparing baseline, fixed-weight, and trainable-weight initialization methods.
- Significant improvements in item coverage were observed, particularly in sequence models, enabling better learning of interest sequences and recommending diverse rather than popularity-biased content.

Experiments

RQ3: Mitigation of Structural Item Bias via Conformity Understanding

- The popularity of recommended items was compared across different user groups.
- Tailored recommendations were identified: popular content increased by 2.3% for users preferring popular items, while popular content was reduced by 3.2% for low-conformity users.
- Most importantly, we show that structural exposure bias was effectively reduced for different user groups.

Conclusion

Summary

- CIPHER is a practical, causal inference-based plug-in framework that addresses structural exposure bias without altering existing recommendation models.
- Validated on real-world IPTV data, it enhances recommendation diversity, provides interpretable causal embeddings, and deepens understanding of true user interests and conformity-driven behaviors, demonstrating substantial practical value for robust personalization.

Future Work

- Offline performance validation of the framework will be conducted in real-world services, extending user understanding to additional observable factors such as price sensitivity and recency.
- Application of intent understanding technology will be expanded beyond media services to other areas within the company.

THANK YOU!