SOVEREIGN: To what extent does the expandable side-MoE architecture improve robustness to distribution shift in user-item

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 29, 2026

Abstract

Streaming recommender systems (SRSs) are widely deployed in real-world applications, where user interests shift and new items arrive over time. As a result, effectively capturing users' latest preferences is challenging, as interactions reflecting recent interests are limited and new items often lack sufficient feedback. A common solution is to enrich item representations using multimodal encoders (e.g., BERT or ViT) to extract visual and textual features. However, these encoders are pretrained on general-purpose tasks: they are not tailored to user preference modeling, and they overlook the f

1 Introduction

Analysis of: Efficient Multimodal Streaming Recommendation via Expandable Side Mixture-of-Experts. Research goal: To what extent does the expandable side-MoE architecture improve robustness to distribution shift in user-item interactions compared to standard continual learning methods (e.g., EWC, SI) on multimodal sequential recommendation datasets like Yelp and Steam?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score $<$ 6.5.

3 Results

12 papers retrieved. 7 claims extracted, 0 verified. Tribunal: 2.8/10 → REJECT (revision_round=0). Policy: ESCALATE_TO_OWNER.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

5 Extracted Claims

Claim	Verified	Confidence
The number of parameters of XSMoE is (2' + +) ≈ O(').	×	0.02
The per-epoch training time complexity of XSMoE is O(').	×	0.03
At any point of the training stage, only one expert, along with the router, remains trainable per layer in XSMoE.	×	0.03
The number of trainable parameters in XSMoE is (2' + +) ≈ O(').	×	0.02
Each expert in XSMoE contains one up-projection layer and one down-projection layer, both of which have ' parameters.	×	0.02
The router in XSMoE has + parameters where is the input size and is the number of experts at each layer.	×	0.04
In XSMoE, the GPU memory usage is dominated by model weights, gradients, optimizer states, and activations.	×	0.03

References

- http://arxiv.org/abs/2002.00741v1
- http://arxiv.org/abs/2312.04693v3
- http://arxiv.org/abs/2508.05993v3