Pre-Trained Recommender Models via Universal User and Item Representations, Evaluating ChatGPT as a Zero-Shot Recommender System, and More!

Vol.16 for Sep 04 - Sep 10, 2023

SUMIT

8 СЕНТ. 2023 Г.

Stay Ahead of the Curve with the Latest Advancements in Information Retrieval and Discoveries.

This week's newsletter highlights the following research:

1. Accelerating Large-Scale Recommendation with Quantized Graph Neural Networks, from Visa Research
2. Capturing Interaction Heterogeneity for Accurate Multi-Behavior Recommendation, from HKU
3. Pre-Trained Recommender Models via Universal User and Item Representations, from UIUC
4. Enhancing Robustness in Recommender Systems: A Survey, from UCAS
5. Joint Explicit and Implicit Scenario Modeling for Fine-Grained Recommendation, from Huawei
6. Balanced and Effective Feature Selection with Multi-View Importance Modeling, from Naver
7. An Industrial-Friendly Framework for Spatial-Temporal Graph Learning in POI Recommendation, from Meituan
8. Modeling User Behavior Hierarchies for Improved Recommendations, from Amazon
9. Optimizing Long-Term Engagement in Voice Shopping with Extended Conversions, from Amazon
10. Evaluating ChatGPT as a Zero-Shot Recommender System, from Poliba

[1] Hessian-aware Quantized Node Embeddings for Recommendation

This paper from Visa Research presents HQ-GNN, a solution for optimizing recommender systems that use Graph Neural Networks (GNNs). GNNs have shown excellent performance but suffer from high latency due to large memory requirements when handling extensive item databases. HQ-GNN addresses this by introducing a quantization module that compresses user/item embeddings into low-bit representations, reducing memory demands and accelerating inference. Unlike previous methods that often use 1-bit quantization, HQ-GNN allows for flexible bit quantization to balance latency and performance. To tackle the gradient mismatch problem associated with quantization, the paper introduces a Generalized Straight-Through Estimator (STE) for better stability and accuracy during training.

📚 https://arxiv.org/abs/2309.01032

[2] Multi-Relational Contrastive Learning for Recommendation

This paper from HKU introduces the Relation-aware Contrastive Learning (RCL) framework, designed to enhance personalized recommender systems. These systems often struggle to capture the diverse ways users interact with items in real-world scenarios, such as clicking, tagging, reviewing, and purchasing. RCL addresses this limitation by incorporating a multi-relational graph encoder to model short-term user interests and a dynamic cross-relational memory network to capture long-term user behavior dependencies. It also employs a multi-relational contrastive learning paradigm to balance commonality and diversity in user behavior representations.

📚 https://arxiv.org/abs/2309.01103

👨🏽‍💻 https://github.com/HKUDS/RCL

[3] Pre-trained Neural Recommenders: A Transferable Zero-Shot Framework for Recommendation Systems

This paper from UIUC explores the concept of pre-trained recommender models (PRMs) for enhancing the efficiency of neural collaborative filtering (NCF) techniques in various domains such as e-commerce and social media. Unlike existing NCF models that require training from scratch for each new domain or dataset, PRMs aim to generalize across different domains with minimal or no retraining and without relying on auxiliary user or item information. The key insight is that the statistical characteristics of user-item interaction data, including user and item marginals and distribution properties, can be used to create universal representations for users and items. These universal features allow for zero-shot adaptation and significantly improve recommendation performance.

📚 https://arxiv.org/abs/2309.01188

https://anonymous.4open.science/r/uni_recsys-2BFE/README.md

[4] Robust Recommender System: A Survey and Future Directions

This paper from UCAS provides a comprehensive survey of the field of robust recommender systems, focusing on strategies to enhance recommender system performance in the face of "dirty" data, which includes malicious attacks and natural noise. It introduces a taxonomy to categorize techniques for improving robustness, such as fraudster detection, adversarial training, certifiable robust training, regularization, purification, and self-supervised learning. The survey covers evaluation metrics and datasets commonly used to assess robustness and discusses how robustness interacts with other properties like accuracy, interpretability, privacy, and fairness in recommender systems. The paper also highlights the growing importance of robustness in recommender systems and outlines open issues and future research directions in this area.

https://arxiv.org/abs/2309.02057

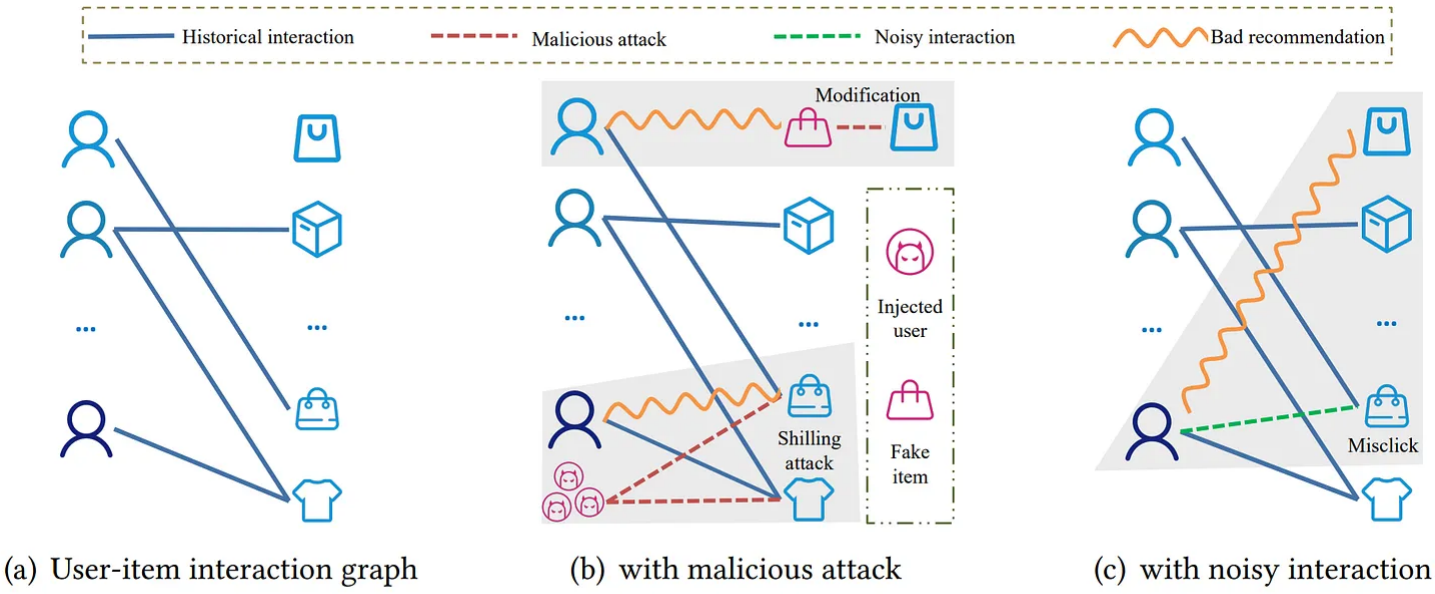

Fig. 2. User-item interaction graph with malicious attack and natural noise.

[5] Scenario-Aware Hierarchical Dynamic Network for Multi-Scenario Recommendation

This paper from Huawei introduces HierRec, a Scenario-Aware Hierarchical Dynamic Network for Multi-Scenario Recommendations, which addresses the challenge of improving Click-Through Rate (CTR) prediction in recommendation and advertising systems by considering both explicit and implicit scenario modeling. Existing multi-scenario models primarily focus on explicit scenario identification based on manual rules, which is sub-optimal. HierRec employs a hierarchical structure with explicit and implicit scenario-oriented layers, using dynamic weights

to capture scenario-specific information. The multi-head implicit modeling design allows for the perception of distinctive patterns, enabling fine-grained scenario modeling.

📚 https://arxiv.org/abs/2309.02061

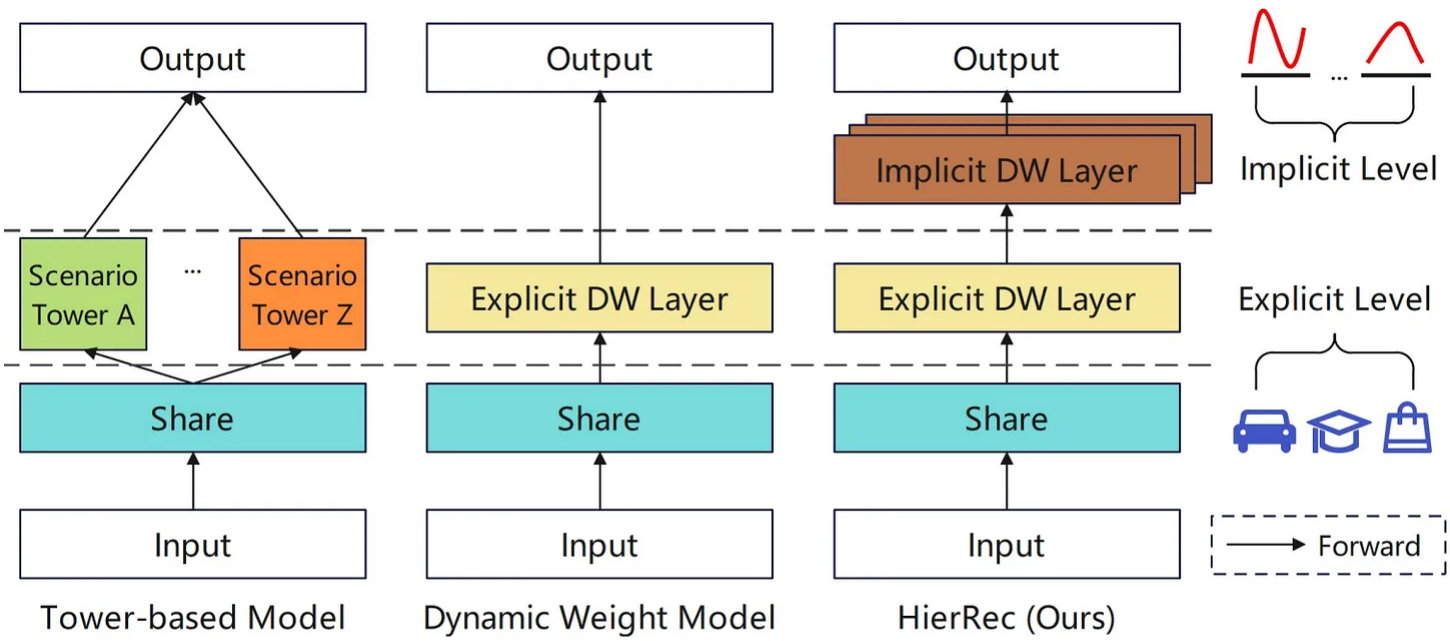

Figure 1: Comparison of different multi-scenario models.

[6] MvFS: Multi-view Feature Selection for Recommender System

This paper from Naver introduces Multi-view Feature Selection (MvFS), a novel approach for feature selection in recommender systems. Traditional feature selection methods often overlook the intricate relationships between features, leading to suboptimal results. MvFS employs a multi-view network consisting of multiple sub-networks, each specializing in different feature patterns to mitigate bias towards dominant patterns and promote a more balanced feature selection process. It also adopts an importance score modeling strategy applied independently to each feature field, avoiding unnecessary dependencies among features.

📚 https://arxiv.org/abs/2309.02064

👨‍💻 https://github.com/dudwns511/MvFS_CIKM23

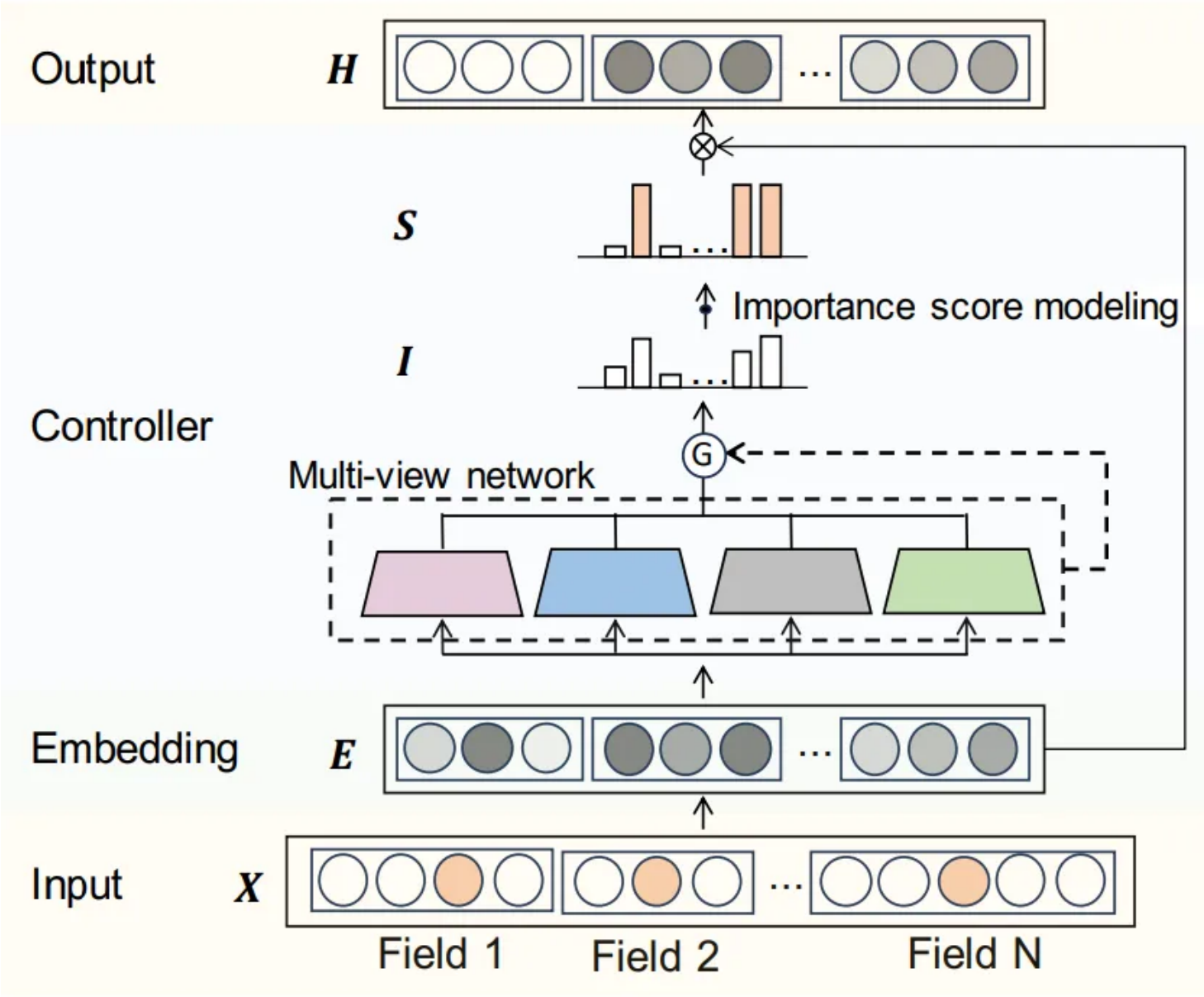

Figure 1: Overview of MvFS.

[7] STGIN: Spatial-Temporal Graph Interaction Network for Large-scale POI Recommendation

This paper from Meituan presents a novel approach for Point-Of-Interest (POI) recommendation in Location-Based Services. Traditional recommendation systems face challenges in capturing user interests across various spatial-temporal contexts. To address this, the authors introduce the Spatial-Temporal Graph Interaction Network (STGIN), which constructs subgraphs representing spatial, temporal, spatial-temporal, and global views to precisely characterize user interests in different contexts. Additionally, an industry-friendly framework is devised to track a user's latest interests in real-time.

https://arxiv.org/abs/2309.02251

[8] Impression-Informed Multi-Behavior Recommender System: A Hierarchical Graph Attention Approach

This paper from Amazon addresses the complexity of multi-behavior interactions in recommender systems, which often simplify diverse user-item interactions into a single 'interaction' label or prioritize only the target behavior, such as 'buy.' To overcome these limitations, the authors introduce the Hierarchical Multi-behavior Graph Attention Network (HMGN). This framework employs attention mechanisms to handle both inter and intra-behaviors and utilizes a multi-task Hierarchical Bayesian Personalized Ranking (HBPR) for optimization. A specialized multi-behavior sub-graph sampling technique is also integrated for scalability.

📚 https://arxiv.org/abs/2309.03169

[9] Extended conversion: Capturing successful interactions in voice shopping

This paper from Amazon addresses the challenge of measuring successful interactions in voice product search, a domain where behavioral signals are limited compared to traditional web shopping. They propose a metric called Extended ConVersion (ECVR) that extends the notion of conversion beyond immediate purchase actions. ECVR considers purchase actions that occur at a later stage during the same shopping journey, even on different channels than where the interaction started. The authors argue that users' behavioral signals should encompass their entire shopping journey, accounting for the fact that voice shopping often starts with informational or exploratory needs and may progress to transactional needs later.

📚 https://www.amazon.science/publications/extended-conversion-capturing-successful-interactions-in-voice-shopping

[10] Evaluating ChatGPT as a Recommender System: A Rigorous Approach

This paper from Poliba explores the potential of ChatGPT as a zero-shot recommender system, conducting comprehensive experiments across three datasets: MovieLens, Last.FM, and Facebook Book. The study assesses ChatGPT's ability to provide recommendations, reorder existing lists, leverage user preferences, handle cold-start situations, and address bias. Results show that ChatGPT can effectively recommend items, with performance comparable to state-of-the-art recommender systems and other large language models. It exhibits characteristics of a hybrid recommender system, leaning towards collaborative and content-based approaches, and displays domain-specific knowledge in books and movies domains. Furthermore, ChatGPT excels in addressing cold-start problems and personalizing recommendations based on user profiles.

📚 https://arxiv.org/abs/2309.03613

👨🏽‍💻 https://github.com/sisinflab/Recommender-ChatGPT

I hope this weekly roundup of top papers has provided you with valuable insights and a glimpse into the exciting advancements taking place in the field. Remember to look deeper into the papers that pique your interest.

I also blog about Machine Learning, Deep Learning, MLOps, and Software Engineering domains. I explore diverse topics, such as Natural Language Processing, Large Language Models, Recommendation Systems, etc., and conduct in-depth analyses, drawing insights from the latest research papers.

5 Likes · 1 Restack

Comments

Write a comment...