Persona Aware Session-Based Recommendation with LLM-Enhanced Knowledge Graphs

Muskan Gupta
muskan.gupta4@tcs.com
TCS Research, India

Priyanka Gupta
priyanka.g35@tcs.com
TCS Research, India

Jyotsana Khatri
khatri.jyotsana@tcs.com
TCS Research, India

Abstract

Session-based recommendation systems (SBRS) effectively capture user's short-term intent from interaction sequences, yet their reliance on session level interactions often limits the modeling of stable long-term user preferences. While recent LLM-augmented recommender systems enrich item representations, effective personalization beyond session signals remains challenging, particularly in sparse-data settings where explicit user profiles are unavailable. We propose a persona-driven session-based recommendation framework that integrates latent user personas inferred from a heterogeneous knowledge graph (KG) into a data-driven SBRS. Our approach follows a two-stage pipeline. First, in the personalized information extraction stage we construct a heterogeneous KG combining user-item interactions, item-item relations, item-feature associations, and external metadata from DBpedia. Using LLM-derived item embeddings as initialization for item nodes, we learn latent user personas in an unsupervised manner via a Heterogeneous Deep Graph Infomax objective. Second , in the personalized information utilization stage, persona representations are incorporated into a modified SBRS for personalized candidate generation, followed by re-ranking with a sequential model to preserve short-term session intent. Experiments on the Amazon Books and Amazon Movies & TV datasets show consistent performance gains across standard ranking metrics compared to baseline methods. These results demonstrate the effectiveness of integrating LLM-enhanced heterogeneous knowledge graphs with latent persona modeling for session-based recommendation.

Keywords

Session-based Recommendation, Heterogeneous Knowledge Graphs, Deep Graph Infomax, Large Language Models, Persona Modeling

ACM Reference Format:
Muskan Gupta, Priyanka Gupta, and Jyotsana Khatri. 2026. Persona Aware Session-Based Recommendation with LLM-Enhanced Knowledge Graphs. In *Proceedings of 3rd Workshop on Evaluating and Applying Recommender Systems with Large Language Models (EARL'26).* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
EARL'26, Rome,Italy
© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Session-based recommendation systems (SBRS) have become a dominant paradigm for capturing short-term user intent through interaction sequences, especially in environments characterized by sparse, unreliable, or non-existent long-term historical data. By emphasizing on recent item sequences, SBRS methods capture transient preferences such as immediate goals and contextual interests. However, a fundamental limitation of most existing SBRS approaches is their emphasis on session level interactions while under utilizing stable user preference signals accumulated across historical interactions. As a result, it constrains personalization especially in sparse data, and cold-start scenarios, or when user behavior is distributed across disjoint sessions. Recent work has attempted to enrich SBRS with side information or user representations, yet the majority of approaches remain predominantly sequence-centric [19, 20, 22, 52, 54]. As a result, they struggle to disentangle short-term intent from longer-term preference signals. In practice, user behavior is shaped not only by immediate session context but also by latent user personas-relatively stable patterns such as brand affinity, aspect-level preferences, stylistic things, or domain-specific inclinations. Explicitly modeling such personas is essential for robust personalization, but is relatively under-explored in SBRS. Effective SBRS requires a mechanism to integrate long-term user preferences with short-term session dynamics, in a way that is efficient, interpretable, and robust across sparse data conditions.

In this work, we propose a persona-aware SBRS framework that incorporates heterogeneous knowledge graphs (KG) and LLM-derived representations for personalized recommendation. Our key insight is that latent user personas represents stable, long-term user preferences which can be inferred from a rich heterogeneous KG that encodes multi-relational links between users, items, item attributes, and external knowledge sources such as DBpedia. These structured relationships capture stable patterns such as genre affinity, stylistic preferences, or topical interests, which pure sequence models often overlook. To this end, we construct a heterogeneous KG integrating three types of information: (i) time-independent user–item interactions, (ii) item–item relational structure, and (iii) item–feature associations drawn from metadata sources such as DBpedia [13]. To learn the latent user persona representations, we first initialize the item nodes and item meta-data nodes in the KG with LLM-derived embeddings and then obtain user persona using a Heterogeneous Deep Graph Infomax (HDGI) objective [25], which captures high-level structural semantic information in an unsupervised manner. Next, we utilize the generated user persona along with LLM-based item representation within SBRS (data-driven). Persona embeddings enrich the retriever with explicit preference cues, while LLM-based item representations capture fine-grained

semantics essential for high-quality candidate generation. To further refine predictions, we incorporate reranking using the base SBRS to provide benefit of short-term intent. Our contributions are: **(1) Persona-aware SBRS**, which combines session intent with KG-derived long-term preferences; **(2) KG-grounded user persona modeling**, which infers user personas using HDGI over a heterogeneous KG; and **(3) Hybrid KG-LLM retrieval**, which integrates KG-derived user personas and LLM-derived item embeddings for recommendation. Unlike prior SBRS methods that rely solely on sequential behavior or text-derived user representations, our approach infers user personas from structured relational KG and integrates them with LLM-derived item semantics. By unifying heterogeneous KG with LLMs, it closes the gap between short-term intent modeling and long-term preference understanding, offering a robust solution for modern recommendation environments.

2 Related Work

Session-based Recommendation SBRS infer user preferences from interaction sequences, where long-term user profiles are limited or unavailable. Early approaches relied on Markov chain [8, 11] and collaborative filtering based models [4, 37]. With the advancements of deep learning, models such as GRU4Rec [9], Caser [28] , SASRec [12] , SRGNN [48] and BERT4Rec [27] have achieved significant improvement by leveraging recurrent neural networks, self-attention mechanism [30] and graph-structured representations [5–7, 23, 39]. More recent graph based approaches such as GCE-GNN [44], LESSR [3], and TAGNN [51], further improve session representation learning by modeling item transitions and target-aware dependencies within sessions.

Moreover, several studies have explored personalized session recommendation by incorporating user aware representations into SBRS [46, 49]. For example, SSE-PT [46] incorporates personalized user embeddings into transformer based recommendation, while [49] leverage user aware graph representations. Despite these advances, these methods learn user representations directly from interaction histories and largely overlook structured relational knowledge.

KG enhanced Recommendation KG based recommendations enrich the sparse user-item interactions by incorporating structured relations among items, item metadata, and attributes, particularly benefiting cold-start scenarios [50]. Representative approaches include RippleNet [33], KGCN [35], KGAT [40], and MKR [34], which leverage KG reasoning, graph message passing, and joint KG-recommendation learning. Furthermore, KG-based approaches mitigate hallucinations and improve generalization and interpretability in LLM-based tasks[1]. In parallel, DGI [31] and its heterogeneous extension HDGI [25] learn node representations in an unsupervised manner by maximizing local-global mutual information, offering a scalable way to derive rich embeddings from heterogeneous structures without labels.

LLM augmented Recommendation LLMs have become increasingly important in recommendation systems due to their semantic understanding and reasoning capabilities [18, 22, 24, 36, 41]. Existing work applies LLMs to feature enrichment [22, 26], candidate generation [10], ranking [42], and explanation [17, 47], often by encoding item metadata into dense semantic representations [14, 16]. Other approaches infer user preferences through history summarization [2], pseudo-profile generation [32, 45], reasoning over past interactions [21], or chain-of-thought prompting [19]. However, these methods typically depend on online LLM inference and may produce personas lacking structural grounding and interpretability. Recent studies have explored combining KG and LLMs for recommendation [15, 38, 53], yet most rely on long user histories and are not designed for session-based settings. In contrast, we learn user personas from a heterogeneous KG enabling user representations to capture semantic and relational preference signals beyond behavioral interaction histories, using HDGI with LLM-derived item embeddings and integrate them into SBRS for personalized candidate generation and session aware re-ranking.

3 Problem Formulation

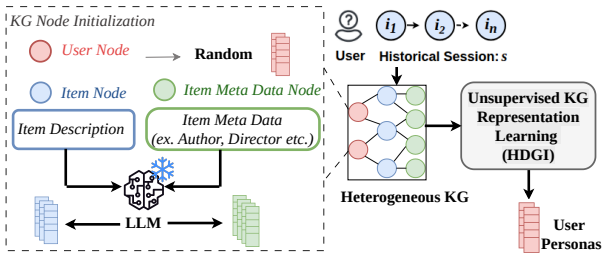
Suppose that $\mathcal{S}$ denotes the set of all sessions containing user-item interactions (e.g. click/rated), and $\mathcal{I}$ denotes the set of n items observed in $\mathcal{S}$. A session $s \in \mathcal{S}$ is represented as an ordered sequence of interactions $s = (i_{s,1}, i_{s,2}, \ldots, i_{s,|s|})$, where $i_{s,j}$ $(j = 1 \ldots |s|) \in \mathcal{I}$, denotes the j^{th} item in session s. The objective of SBRS is to predict the next item $i_{s,|s|+1}$ by estimating an n-dimensional relevance vector $\hat{\mathbf{y}}_{s,|s|+1}$. The top-$K$ highest-scoring items are returned as recommendations.

Heterogeneous KG: A heterogeneous KG is a multi-relational graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \phi, \psi)$ in which each node $v \in \mathcal{V}$ and each edge $e \in \mathcal{E}$ has a type specified by the mappings $\phi : \mathcal{V} \rightarrow \mathcal{T}_V$ (node types) and $\psi : \mathcal{E} \rightarrow \mathcal{T}_E$ (edge types). A graph is considered heterogeneous if it contains multiple node or edge types, i.e.,$|\mathcal{T}_V| + |\mathcal{T}_E| \geq 2$. Nodes may contain attribute information represented by an initial feature matrix: $\mathbf{X}$. *Heterogeneous Graph Representation Learning*: Given $\mathcal{G}$ and $\mathbf{X}$, the goal of heterogeneous representation learning task is to learn the low dimensional node representations $\mathcal{H} \in \mathcal{R}^{|\mathcal{V}| \times d}$ which contains both structure information from graph $\mathcal{G}$ and semantic information encoded in $\mathbf{X}$.

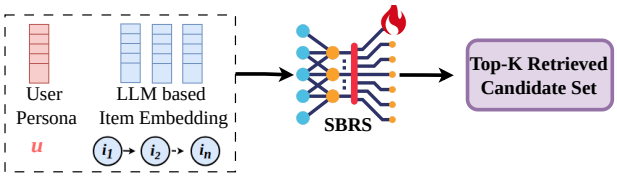
4 Proposed Approach

In this work, our goal is to effectively leverage KG-grounded user personas/preferences and rich item semantics to improve SBRS under settings where long-term user preferences are only available through historical interactions The input consists of a set of sessions $\mathcal{S}$, a heterogeneous KG, $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \phi, \psi)$ which integrates time-independent user-item interactions, item-item relations (e.g., also-buy, also-view), and item-feature associations (e.g. brands, category), and external semantic metadata derived from DBPedia [13]). Using this graph, we infer a KG-grounded user personas that capture latent user personas. Additionally, we derive semantic item embeddings using large language models (LLMs) from textual signals such as product descriptions and metadata.

Inspired by CoTRec [19], as shown in Figure 1, our framework consists of two stages, (i) *Stage-1 Personalized Information Extraction*, which constructs an LLM-initialized heterogeneous KG and learns KG-grounded user personas through unsupervised representation learning and (ii) *Stage-2 Personalized Information Utilization*, which leverages the extracted persona and LLM based semantic signals in a data-driven SBRS and rerank it using a base SBRS Model. The framework uses KG-driven persona modeling to capture user personas beyond short-term session signals, while using the base sequential model as the reranker.



(a) Stage-1 Personalized Information Extraction



(b) Stage-2 Personalized Information Utilization

Figure 1: Overview of persona-aware SBRS framework.

Stage-1 Personalized Information Extraction: Existing methods primarily rely on item interaction sequences to infer user intent. We leverage heterogeneous KG to infer user personas that reflect stable, long-term user preferences derived from semantic and relational information. We construct a hybrid KG that integrates time independent user-item interaction signals and item meta data form external knowledge sources such as DBpedia [13] and items viewed/rated by same user. We utilize Amazon-KG [43][1] which enriches Amazon interaction data with structured information derived from DBPedia. This hybrid KG encodes time-independent user preferences, complementing session-based interaction sequences that primarily capture short-term intent. To incorporate semantic information, item and metadata nodes are initialized with LLM-derived embeddings obtained from textual descriptions. This injects semantic knowledge into the heterogeneous KG and aligns different node types within a shared embedding space. Since users do not possess textual profile, descriptions, user embeddings are randomly initialized and learned through graph-level self-supervision. To capture rich and semantically grounded user preferences from the heterogeneous KG, we employ an unsupervised representation learning framework based on HDGI [25]. Following [25], the heterogeneous KG consists of multiple node types (e.g., users, items, attributes, entities) and relation types. HDGI utilizes meta-path structures to capture semantic connectivity patterns across heterogeneous nodes. For a given relation type r, the neighborhood aggregation is performed through a graph convolution module specialized for heterogeneous graphs. After HDGI training, we directly extract the embeddings of the user nodes from the shared heterogeneous representation space to represent user personas. Because HDGI incorporates meta-path semantics, graph convolution aggregation, and mutual-information-based contrastive learning, the resulting embeddings encode multi-hop, multi-relational preference signals.

Stage-2 Personalized Information Utilization: In this stage, we utilize the KG-aware User Personas and the LLM based Item embeddings generated in Stage-1, to retrieve the top-K candidate items using data-driven SBRS. Inspired by the personalized information utilization framework of CoT-Rec [19], we inject the personalized information into a data-driven SBRS model for candidate generation. However, unlike CoT-Rec, which utilizes personalized information within an LLM-powered recommendation framework, we incorporate KG-aware user personas directly into SBRS. To introduce long-term preference information into session modeling, the KG-aware user persona representation is incorporated as an additional input to the SBRS architecture together with the item embeddings in the session. By jointly encoding persona information and session interactions, the model learns representations that capture both stable user preferences and sequential behavioral dependencies. The retrieved candidate items are further refined using the base SBRS as the reranker to focus more on short-term intent. Given a session sequence the sequential model encodes the session context and computes relevance scores for candidate items. The model outputs relevance scores over the candidate set and the top-K items are returned as the final recommendations. By combining KG-derived persona embeddings with SBRS, the proposed framework enables the retriever to utilize both short-term behavioral signals and long-term preference information encoded in the KG.

[1]https://github.com/WangYuhan-0520/Amazon-KG-v2.0-dataset

Table 1: Statistics of the datasets used for experiments.

Datasets	#users	#items	#interactions	#entities	#triples
Amazon-Movies & TV	10,849	3,672	87,376	45,528	209,630
Amazon-Books	1,624	1,345	10,947	8,660	22,830

5 Experimental Evaluation

Dataset and Evaluation Metrics: We evaluate our framework on two publicly available benchmark datasets: Amazon Movies & TV and Amazon Books from the Amazon Reviews [2]. Following prior session-based recommendation work, interactions are chronologically ordered and users/items with fewer than five interactions are removed. These datasets contain user-item interaction sequences and rich item metadata such as titles, categories, and descriptions. The statistics of datasets are provided in Table 1. We report standard top-K recommendation metrics including Hit Rate (HR@K), Normalized Discounted Cumulative Gain (NDCG@K), and Mean Reciprocal Rank (MRR).

To model long-term user preferences, we construct the heterogeneous KG, using Amazon-KG [3], which enriches interaction data with structured item attributes and entity relations derived from DBpedia. The graph contains user, item, and metadata entities connected through multiple semantic relations. In our heterogeneous KG, each user node is connected to the items' nodes that user had interacted with and item nodes are further connected to their corresponding metadata entities (e.g., author, director, genre, subject, category). Unlike user nodes, item and metadata nodes possess rich textual information that conveys semantic meaning. During training, semantic information from LLM-initialized item and metadata nodes propagates through the graph structure to user nodes via user-item and item-metadata connections. As a result, the learned user embeddings capture both the interaction behavior of users and the semantic characteristics of the items they engage with. This design combines the complementary strengths of LLMs and graph representation learning: LLMs provide semantically meaningful

[2]https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/
[3]https://github.com/WangYuhan-0520/Amazon-KG-v2.0-dataset

Table 2: Comparison of baseline and proposed methods.

Method	k	User Embedding	Item Embedding	Amazon Movies & TV			Amazon Books		
				HR	NDCG	MRR	HR	NDCG	MRR
Vanilla SASRec	10	None	Random	0.1302	0.0752	0.0585	0.1693	0.1109	0.093
SASRec + LLM Item Embeddings	10	None	Description (LLM)	0.1314	0.0774	0.061	0.1791	0.1075	0.0857
SASRec + Random User Embeddings	10	Random	Description (LLM)	0.1116	0.0624	0.0473	0.1711	0.1005	0.0788
SASRec + LLM User Profile	10	LLM	Description (LLM)	0.1171	0.0648	0.049	0.1785	0.1077	0.086
Proposed: Persona SASRec	10	KG+LLM	Description (LLM)	**0.1327**	0.0764	0.0593	0.1773	0.1101	0.0894
Persona SASRec + ReRanking	10	KG+LLM	Description (LLM)	**0.1327**	**0.0776**	**0.0609**	**0.1773**	**0.1123**	**0.0925**
Vanilla SASRec	100	None	Random	0.3665	0.1220	0.0666	0.4254	0.1607	0.1012
SASRec + LLM Item Embeddings	100	None	Description (LLM)	0.3639	0.1235	0.069	0.427	0.1559	0.0939
SASRec + Random User Embeddings	100	Random	Description (LLM)	0.3507	0.1097	0.0555	0.4347	0.1525	0.0877
SASRec + LLM User Profile	100	LLM	Description (LLM)	0.3527	0.111	0.0568	0.4384	0.1588	0.0948
Proposed: Persona SASRec	100	KG+LLM	Description (LLM)	**0.3697**	**0.1229**	0.0672	**0.4427**	**0.1622**	0.0982
Persona SASRec + ReRanking	100	KG+LLM	Description (LLM)	**0.3697**	**0.1236**	**0.0677**	**0.4427**	**0.1641**	**0.1016**

initial representations for textual entities, while HDGI leverages the heterogeneous graph structure to propagate this information and infer latent user personas from multi-relational interactions.

Baselines and Personalization Strategies: To evaluate the effectiveness we compare several SASRec [12] variants to isolate the contributions of semantic item representations and user personalization. Specifically, we consider: Vanilla SASRec without explicit user modeling; SASRec with LLM-derived item embeddings; SASRec with randomly initialized user embeddings; SASRec with LLM-generated user profiles; Persona-SASRec, which employs HDGI-derived KG personas with LLM item embeddings; and Persona-SASRec+Rerank, which further applies reranking using base SASRec model to better capture short-term session intent.

Hyperparameter Setup: We use Qwen-3-8B [29][4] model wherever an LLM call is required for text generation, and Qwen3-Embedding-8B[5] for generating text embeddings. Our base SBRS model is SASRec [12]. We use validation data for hyperparameter selection using NDCG@10 as the performance metric for all approaches. We tuned the architectural parameters over Transformer depth $L \in \{2, 3, 4\}$, hidden dimensionality $d \in \{64, 128, 256\}$, number of attention heads $\{2, 4\}$, and dropout rates $\{0.1, 0.2, 0.3, 0.5\}$. The results shown in Table 2 correspond to the configuration with $d_t = 64$, dropout = 0.1, learning rate = $1e-3$, and $L = 2$. All models were trained using the Adam optimizer.

5.1 Results and Discussion

Table 2 compares the performance of different user representation strategies used with SBRS. The results demonstrate that incorporating KG-derived user personas consistently improves recommendation performance across both datasets.

The inclusion of KG-based user preference embeddings leads to higher HR@100 across both datasets, this suggests that the heterogeneous KG captures stable long-term user preferences through relational signals, enabling the retriever to generate more relevant candidate items. We further observe that LLM-derived item embeddings often achieve better retrieval performance than randomly initialized item embeddings, particularly in terms of candidate recall. While improvements are not uniform across all datasets and metrics, the results suggest that LLM-generated descriptions provide complementary semantic signals that help identify related items besides interaction data.

For the Amazon Books dataset, improvements in HR@10 are relatively modest due to the high sparsity of interaction data. The heterogeneous KG introduces additional relational signals that allow the retriever to identify more relevant items, including long-tail items connected through semantic relations (i.e., better HR). However, these items may appear lower in the ranked list (i.e., low NDCG and MRR), making the improvement more visible when the candidate cutoff increases to k=100.

After candidate retrieval, we apply reranking using vanilla SASRec. Although the reranker does not directly use KG-based persona information, it improves MRR and NDCG, which measure the ranking quality within the candidate set. These results indicate that persona-aware retrieval primarily contributes to candidate retrieval and coverage, whereas sequential reranking contributes to ranking quality. This improvement occurs because vanilla SASRec models sequential dependencies within the session, allowing the model to rank better candidate items according to the user's short-term intent. This also supports the motivation of explicitly separating long-term preference modeling from short-term intent modeling in session-based recommendation.

In conclusion, the results demonstrate that combining KG-derived persona representations with LLM-based item semantics yields more effective candidate retrieval than relying on session interactions alone. Although the observed improvements are moderate on certain metrics, the gains are consistent across datasets and retrieval cutoffs, highlighting the complementary nature of structured knowledge, semantic representations, and sequential recommendation models.

6 Conclusion

We presented a persona-aware SBRS framework that combines KG-derived user personas learned with HDGI and LLM-based item semantic representations. Experiments on Amazon Books and Movies & TV demonstrate consistent improvements over sequential baselines, highlighting the benefits of integrating structured knowledge with semantic representations. Future work will investigate stronger persona-session fusion mechanisms and comparisons with additional personalized session-based recommendation baselines.

[4]https://huggingface.co/Qwen/Qwen3-8B
[5]https://huggingface.co/Qwen/Qwen3-Embedding-8B

References

[1] Garima Agrawal, Tharindu Kumarage, Zeyad Alghamdi, and Huan Liu. 2024. Can Knowledge Graphs Reduce Hallucinations in LLMs? : A Survey. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, Kevin Duh, Helena Gomez, and Steven Bethard (Eds.). Association for Computational Linguistics, Mexico City, Mexico, 3947–3960. doi:10.18653/v1/2024.naacl-long.219

[2] Seunghwan Bang and Hwanjun Song. 2025. Llm-based user profile management for recommender system. *arXiv preprint arXiv:2502.14541* (2025).

[3] Tianwen Chen and Raymond Chi-Wing Wong. 2020. Handling information loss of graph neural networks for session-based recommendation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 1172–1180.

[4] Michael D Ekstrand, John T Riedl, Joseph A Konstan, et al. 2011. Collaborative filtering recommender systems. *Foundations and Trends® in Human–Computer Interaction* 4, 2 (2011), 81–173.

[5] Muskan Gupta, Priyanka Gupta, Jyoti Narwariya, Lovekesh Vig, and Gautam Shroff. 2024. SCM4SR: Structural Causal Model-based Data Augmentation for Robust Session-based Recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2609–2613.

[6] Muskan Gupta, Priyanka Gupta, and Lovekesh Vig. 2024. Guided Diffusion-based Counterfactual Augmentation for Robust Session-based Recommendation. *arXiv preprint arXiv:2410.21892* (2024).

[7] Priyanka Gupta, Diksha Garg, Pankaj Malhotra, Lovekesh Vig, and Gautam Shroff. 2019. NISER: Normalized Item and Session Representations with Graph Neural Networks. *arXiv preprint arXiv:1909.04276* (2019).

[8] Ruining He and Julian McAuley. 2016. Fusing similarity models with markov chains for sparse sequential recommendation. In *2016 IEEE 16th international conference on data mining (ICDM)*. IEEE, 191–200.

[9] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2015. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939* (2015).

[10] Jiani Huang, Shijie Wang, Liangbo Ning, Wenqi Fan, Shuaiqiang Wang, Dawei Yin, and Qing Li. 2026. Towards next-generation recommender systems: A benchmark for personalized recommendation assistant with llms. In *Proceedings of the Nineteenth ACM International Conference on Web Search and Data Mining*. 217–226.

[11] Mohsen Jamali and Martin Ester. 2010. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the fourth ACM conference on Recommender systems*. 135–142.

[12] Wang-Cheng Kang and Julian McAuley. 2018. Self-Attentive Sequential Recommendation. In *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 197–206. doi:10.1109/ICDM.2018.00035

[13] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. 2015. DBpedia — A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. *Semantic Web* 6, 2 (2015), 167–195. doi:10.3233/SW-140134

[14] Jiacheng Li, Ming Wang, Jin Li, Jinmiao Fu, Xin Shen, Jingbo Shang, and Julian McAuley. 2023. Text is all you need: Learning language representations for sequential recommendation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1258–1267.

[15] Siqi Liang, Yudi Zhang, and Yue Guo. 2025. PersonaAgent with GraphRAG: Community-Aware Knowledge Graphs for Personalized LLM. *arXiv preprint arXiv:2511.17467* (2025). https://arxiv.org/abs/2511.17467

[16] Jan Malte Lichtenberg, Antonio De Candia, and Matteo Ruffini. 2025. DenseRec: Revisiting Dense Content Embeddings for Sequential Transformer-based Recommendation. *arXiv preprint arXiv:2508.18442* (2025).

[17] Jianghao Lin, Xinyi Dai, Yunjia Xi, Weiwen Liu, Bo Chen, Hao Zhang, Yong Liu, Chuhan Wu, Xiangyang Li, Chenxu Zhu, et al. 2025. How can recommender systems benefit from large language models: A survey. *ACM Transactions on Information Systems* 43, 2 (2025), 1–47.

[18] Fan Liu, Yaqi Liu, Huilin Chen, Zhiyong Cheng, Liqiang Nie, and Mohan Kankanhalli. 2025. Understanding before recommendation: Semantic aspect-aware review exploitation via large language models. *ACM Transactions on Information Systems* 43, 2 (2025), 1–26.

[19] Jiahao Liu, Xueshuo Yan, Dongsheng Li, Guangping Zhang, Hansu Gu, Peng Zhang, Tun Lu, Li Shang, and Ning Gu. 2025. Improving llm-powered recommendations with personalized information. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2560–2565.

[20] Xin Liu, Zheng Li, Yifan Gao, Jingfeng Yang, Tianyu Cao, Zhengyang Wang, Bing Yin, and Yangqiu Song. 2023. Enhancing user intent capture in session-based recommendation with attribute patterns. *Advances in Neural Information Processing Systems* 36 (2023), 30821–30839.

[21] Hanjia Lyu, Song Jiang, Hanqing Zeng, Yinglong Xia, Qifan Wang, Si Zhang, Ren Chen, Chris Leung, Jiajie Tang, and Jiebo Luo. 2024. Llm-rec: Personalized recommendation via prompting large language models. In *Findings of the Association for Computational Linguistics: NAACL 2024*. 583–612.

[22] Jyoti Narwariya, Priyanka Gupta, Muskan Gupta, Jyotsana Khatri, and Lovekesh Vig. 2025. SemSR: Semantics aware robust Session-based Recommendations. *arXiv preprint arXiv:2508.20587* (2025).

[23] Shutong Qiao, Wei Zhou, Junhao Wen, Hongyu Zhang, and Min Gao. 2023. Bi-channel multiple sparse graph attention networks for session-based recommendation. In *Proceedings of the 32nd ACM international conference on information and knowledge management*. 2075–2084.

[24] Xubin Ren, Wei Wei, Lianghao Xia, Lixin Su, Suqi Cheng, Junfeng Wang, Dawei Yin, and Chao Huang. 2024. Representation learning with large language models for recommendation. In *Proceedings of the ACM web conference 2024*. 3464–3475.

[25] Yuxiang Ren, Bo Liu, Chao Huang, Peng Dai, Liefeng Bo, and Jiawei Zhang. 2019. Heterogeneous deep graph infomax. *arXiv preprint arXiv:1911.08538* (2019).

[26] Kainan Shi, Peilin Zhou, Ge Wang, Han Ding, and Fei Wang. 2025. What Matters in LLM-Based Feature Extractor for Recommender? A Systematic Analysis of Prompts, Models, and Adaptation. *arXiv preprint arXiv:2509.14979* (2025).

[27] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM'19)*. ACM, Beijing, China, 1441–1450. doi:10.1145/3357384.3357895

[28] Jiaxi Tang and Ke Wang. 2018. Personalized Top-N Sequential Recommendation via Convolutional Sequence Embedding. In *Proceedings of the 11th ACM International Conference on Web Search and Data Mining (WSDM'18)*. ACM, Marina del Rey, CA, USA, 565–573. doi:10.1145/3159652.3159656

[29] Qwen Team. 2025. Qwen3 Technical Report. arXiv:2505.09388 [cs.CL] https://arxiv.org/abs/2505.09388

[30] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[31] Petar Veličković, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R. Devon Hjelm. 2019. Deep Graph Infomax. In *International Conference on Learning Representations (ICLR) — Poster*. https://arxiv.org/abs/1809.10341

[32] Haoting Wang, Haokai Lu, Zheyun Feng, Jenny Huang, Yifat Amir, Gregory Hinkson, Ben Most, Zelong Zhao, Yixin Kelly Cui, Rein Zhang, et al. 2026. LLM-Based User Personas for Recommendations at Scale. *arXiv preprint arXiv:2606.12198* (2026).

[33] Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2018. RippleNet: Propagating User Preferences on the Knowledge Graph for Recommender Systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM'18)*. ACM, Torino, Italy, 417–426. doi:10.1145/3269206.3271739

[34] Hongwei Wang, Fuzheng Zhang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2019. Multi-task feature learning for knowledge graph enhanced recommendation. In *The world wide web conference*. 2000–2010.

[35] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. 2019. Knowledge Graph Convolutional Networks for Recommender Systems. In *The World Wide Web Conference (WWW'19)*. ACM, San Francisco, CA, USA, 3307–3313. doi:10.1145/3308558.3313417

[36] Lu Wang, Di Zhang, Fangkai Yang, Pu Zhao, Jianfeng Liu, Yuefeng Zhan, Hao Sun, Qingwei Lin, Weiwei Deng, Dongmei Zhang, et al. 2025. Lettingo: Explore user profile generation for recommendation system. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*. 2985–2995.

[37] Meirui Wang, Pengjie Ren, Lei Mei, Zhumin Chen, Jun Ma, and Maarten De Rijke. 2019. A collaborative session-based recommendation approach with parallel memory modules. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*. 345–354.

[38] Shijie Wang, Wenqi Fan, Yue Feng, Lin Shanru, Xinyu Ma, Shuaiqiang Wang, and Dawei Yin. 2025. Knowledge graph retrieval-augmented generation for llm-based recommendation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 27152–27168.

[39] Shoujin Wang, Liang Hu, Yan Wang, Xiangnan He, Quan Z Sheng, Mehmet A Orgun, Longbing Cao, Francesco Ricci, and Philip S Yu. 2021. Graph learning based recommender systems: A review. *arXiv preprint arXiv:2105.06339* (2021).

[40] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. KGAT: Knowledge Graph Attention Network for Recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'19)*. ACM, Anchorage, AK, USA, 950–958. doi:10.1145/3292500.3330989

[41] Yan Wang, Zhixuan Chu, Xin Ouyang, Simeng Wang, Hongyan Hao, Yue Shen, Jinjie Gu, Siqiao Xue, James Zhang, Qing Cui, et al. 2024. Llmrg: Improving recommendations through large language model reasoning graphs. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 38. 19189–19196.

[42] Yaqi Wang, Haojia Sun, and Shuting Zhang. 2025. LLM as Explainable Re-Ranker for Recommendation System. *arXiv preprint arXiv:2512.03439* (2025).

[43] Yuhan Wang, Qing Xie, Mengzi Tang, Lin Li, Jingling Yuan, and Yongjian Liu. 2024. Amazon-kg: A knowledge graph enhanced cross-domain recommendation dataset. In *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval*. 123–130.

[44] Ziyang Wang, Wei Wei, Gao Cong, Xiao-Li Li, Xian-Ling Mao, and Minghui Qiu. 2020. Global context enhanced graph neural networks for session-based recommendation. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*. 169–178.

[45] Stanisław Woźniak, Jacek Duszenko, Jan Kocoń, and Przemysaw Kazienko. 2025. Improving llm-based recommender systems with user-controllable profiles. In *Companion Proceedings of the ACM on Web Conference 2025*. 2102–2111.

[46] Liwei Wu, Shuqing Li, Cho-Jui Hsieh, and James Sharpnack. 2020. SSE-PT: Sequential recommendation via personalized transformer. In *Proceedings of the 14th ACM conference on recommender systems*. 328–337.

[47] Likang Wu, Zhi Zheng, Zhaopeng Qiu, Hao Wang, Hongchao Gu, Tingjia Shen, Chuan Qin, Chen Zhu, Hengshu Zhu, Qi Liu, et al. 2024. A survey on large language models for recommendation. *World Wide Web* 27, 5 (2024), 60.

[48] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-based Recommendation with Graph Neural Networks. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*.

[49] Shu Wu, Mengqi Zhang, Xin Jiang, Ke Xu, and Liang Wang. 2019. Personalized graph neural networks with attention mechanism for session-aware recommendation. *arXiv e-prints* (2019), arXiv–1910.

[50] Wooseong Yang, Weizhi Zhang, Yuqing Liu, Yuwei Han, Yu Wang, Junhyun Lee, and Philip S Yu. 2025. Cold-start recommendation with knowledge-guided retrieval-augmented generation. *arXiv e-prints* (2025), arXiv–2505.

[51] Feng Yu, Yanqiao Zhu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2020. TAGNN: Target attentive graph neural networks for session-based recommendation. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*. 1921–1924.

[52] Xiaokun Zhang, Bo Xu, Chenliang Li, Bowei He, Hongfei Lin, Chen Ma, and Fenglong Ma. 2025. A Survey on Side Information-driven Session-based Recommendation: From a Data-centric Perspective. *IEEE Transactions on Knowledge and Data Engineering* (2025).

[53] Xiaokun Zhang, Bo Xu, Chenliang Li, and Yao Zhou. 2024. Side Information-Driven Session-based Recommendation: A Survey. *arXiv preprint arXiv:2402.17129* (2024).

[54] Xiaokun Zhang, Bo Xu, Fenglong Ma, Zhizheng Wang, Liang Yang, and Hongfei Lin. 2026. Rethinking contrastive learning in session-based recommendation. *Pattern Recognition* 169 (2026), 111924.