

Transformer-Based Multi-Behavior Sequential Recommendation Method

Weining Bai, Bo He^{*}, Xia Yu, Yunfei Chen

College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China

^{*} Corresponding author: (Email: hebo@cqut.edu.cn)

Abstract: Multi-Behavior Sequential Recommendation (MBSR) is an important research direction in sequential recommendation. It aims to exploit auxiliary behaviors in users' historical interactions to improve item prediction under a target behavior. In recent years, Transformer architectures have provided an effective foundation for MBSR because self-attention can model long-range sequential dependencies and interactions among heterogeneous behaviors. This paper reviews recent advances in Transformer-based MBSR. We first introduce the task formulation, research background, and representative application scenarios. Representative methods are then reviewed in terms of behavior representation, relation modeling, temporal modeling, and interest modeling, with emphasis on how auxiliary behaviors are used to improve target-behavior item prediction and characterize evolving user interests. Finally, we summarize the main challenges, with particular attention to noisy auxiliary behaviors, efficient long-sequence modeling, and coordination among multiple target behaviors, and discuss promising directions for future research.

Keywords: Multi-Behavior Sequential Recommendation, Transformer, Sequential Recommendation, Auxiliary Behavior, Interest Modeling.

1. Introduction

As the scale of products, content, and services on online platforms continues to grow, recommender systems have become essential for alleviating information overload and improving information access. Traditional recommendation methods generally characterize a user's overall preference from historical interactions. In real-world settings, however, user interests are not static; they evolve continuously with ongoing interactions. Recent clicks, favorites, or consumption behaviors often reflect current interests, whereas earlier interactions may reveal relatively stable long-term preferences. Sequential recommendation therefore preserves the order of interactions, models the temporal evolution of user interests, and predicts subsequent interactions [1–6].

User feedback on real platforms is also inherently heterogeneous. Different behaviors vary in semantics, frequency, and preference strength, while exhibiting temporal dependencies and transition patterns. In e-commerce, for example, browsing and clicking often indicate preliminary interest, favorites and cart additions may signal stronger purchase intention, and purchases represent relatively explicit decisions. These behaviors should therefore not be treated as equivalent. Collapsing them into a single item sequence can obscure behavior-specific preference signals and fail to describe how user interests strengthen or shift over time. Multi-Behavior Sequential Recommendation (MBSR) extends sequential recommendation by explicitly introducing behavior types and jointly modeling interaction order and behavioral heterogeneity, thereby providing a finer characterization of dynamic preferences and decision processes [7–11].

Transformer-based architectures have become major modeling approaches for MBSR. Their self-attention mechanisms establish dependencies among interactions across an entire sequence and can flexibly incorporate behavior types, temporal information, and other contextual

features. Compared with recurrent models that propagate information step by step, Transformer is better suited to capturing long-range dependencies, allowing earlier interactions to contribute directly to current prediction. In multi-behavior settings, historical interactions can also be weighted according to behavior type and occurrence time, supporting the joint characterization of short-term needs and long-term interests. Moreover, the extensible Transformer architecture can be combined with graph modeling, interest learning, and other representation-learning techniques, making it a flexible foundation for modeling cross-behavior and cross-temporal dependencies [12–16].

2. Background And Overview

2.1. Task Definition

Let the user set, item set, and behavior-type set be denoted by U , I , and B , respectively. For a user u , the chronologically ordered multi-behavior interaction sequence is defined as

$$S_u = [(i_1, b_1, t_1), (i_2, b_2, t_2), \dots, (i_T, b_T, t_T)] \quad (1)$$

Where i_t , b_t , and t_t denote the item, behavior type, and temporal information of the t -th interaction, respectively, and T is the sequence length. Different behaviors play different roles in MBSR. Some serve as auxiliary behaviors that provide complementary evidence about user interests and decision processes, whereas the recommendation objective focuses on item prediction under a specified target behavior. Accordingly, the MBSR task is to exploit the associations between auxiliary and target behaviors in a user's historical multi-behavior sequence, characterize the user's current interest state, and predict the candidate item most likely to be interacted with under the target behavior.

Unlike single-behavior sequential recommendation, which mainly focuses on item order, MBSR must distinguish the preference semantics expressed by different behaviors and assess the actual contribution of auxiliary behaviors to target-

behavior prediction. Unlike general multi-behavior recommendation, MBSR explicitly preserves temporal order and therefore considers not only which behaviors occurred but also how they occurred in sequence. MBSR consequently exhibits two fundamental properties: behavioral

heterogeneity and temporal dependence. Because the contribution of an auxiliary behavior may vary with temporal position, user state, and target task, a central challenge is to exploit auxiliary information effectively while modeling its dynamic relationship with the target behavior.

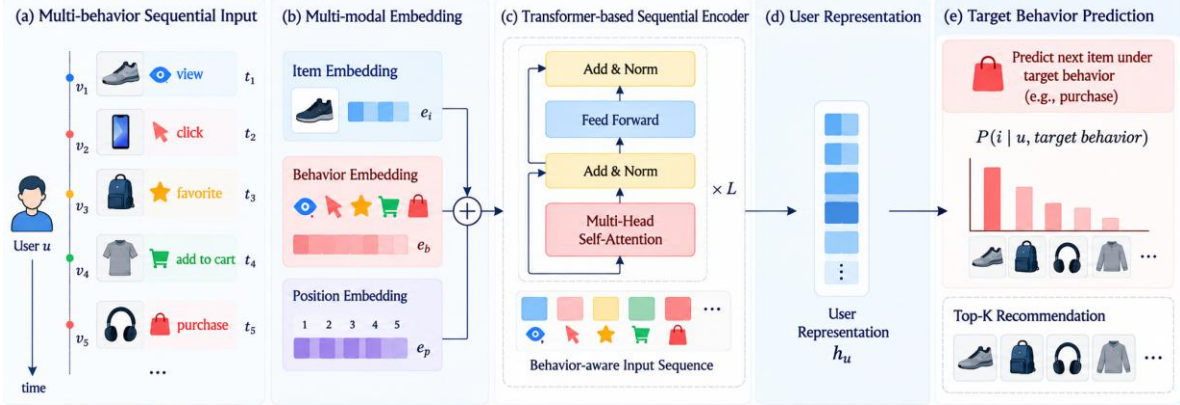


Figure 1. Overall task pipeline of transformer-based multi-behavior sequential recommendation

2.2. Application Scenarios

MBSR is mainly applied to online platforms with diverse interaction types and continuously evolving user preferences. Behavior semantics, behavior relationships, and recommendation objectives vary across application domains, leading to different modeling emphases.

In e-commerce, browsing, clicking, favoriting, adding to cart, and purchasing often correspond to different stages of a decision process. High-frequency auxiliary behaviors provide richer interest cues, whereas target behaviors such as purchase are relatively sparse. Multi-behavior information can therefore complement the limited signals contained in the target behavior alone and help characterize the progression from initial interest to final conversion [9–11].

On short-video, news, music, and social media platforms, behaviors such as viewing, liking, commenting, favoriting, and sharing primarily reflect different levels of content interest and engagement. These behaviors do not necessarily follow a strict one-way conversion path, and users may switch rapidly among topics. Compared with transaction-oriented scenarios, content recommendation places greater emphasis on behavioral semantics, short-term interest shifts, and the continuing evolution of long-term preferences [7–8].

3. Transformer-Based Multi-Behavior Sequential Recommendation

Transformer provides a unified sequence-modeling framework for MBSR, but standard self-attention has limited ability to explicitly characterize behavior-type differences, cross-behavior relations, and dynamic interests. Existing studies therefore extend input representations, attention computation, and user-state modeling to exploit heterogeneous behaviors and their temporal relationships more effectively. In this survey, these methods are organized into four categories: behavior representation, relation modeling, temporal modeling, and interest modeling.

3.1. Behavior Representation

In a multi-behavior sequence, an item alone cannot fully describe the meaning of an interaction. Interactions with the same item but different behavior types may indicate different levels of interest and decision intention. Behavior

representation therefore jointly encodes item and behavior information so that each time step reflects both what the user interacted with and how the interaction occurred. A common formulation is

$$\mathbf{x}_t = \mathbf{e}_t^i + \mathbf{e}_t^b + \mathbf{e}_t^p \quad (2)$$

Where $\mathbf{e}_t^i$, $\mathbf{e}_t^b$, and $\mathbf{e}_t^p$ are the item, behavior, and positional embeddings of the t -th interaction. A temporal embedding $\mathbf{e}_t^t$ can also be introduced when explicit time information is required. This joint representation distinguishes interaction types at the input level. However, fixed behavior embeddings remain limited in capturing user-specific differences and context-dependent behavioral semantics, motivating more flexible joint and personalized representations.

Early Transformer-based methods mainly enhanced sequential representations through behavior embeddings. MB-STR jointly encodes item and behavior information with a Multi-Behavior Transformer and employs a multi-behavior sequential pattern generator to learn ordering patterns among behaviors, allowing behavior information to participate in both sequence modeling and target prediction [13]. MBSRec uses a relatively simple multi-head self-attention architecture to model heterogeneous historical interactions and applies weighted training to regulate the influence of different behaviors [17]. PBAT further considers user-specific behavior patterns and employs a personalized behavior-pattern generator so that the same behavior can be represented differently for different users [16].

As research progressed, behavior representation evolved from simple item-plus-behavior embeddings toward richer interaction-level representations. GEAR organizes item, behavior, and temporal context into interaction units and adopts alternating modeling to learn dependencies among these signals while reducing interference caused by direct feature mixing [18]. BLADE fuses item and behavior information at both the input and intermediate representation levels and combines behavior-level data augmentation with contrastive learning to improve representations of sparse behaviors [19]. BASE further introduces behavior information and item attributes into sequence encoding through sequence, behavior, and attribute pathways, and employs behavior-aware dual-granularity contrastive learning

to strengthen the use of item and attribute information [20].

3.2. Relation Modeling

Interactions in a multi-behavior sequence are not independent; they exhibit transition, co-occurrence, and dependency relationships. A behavior may constitute an important precursor to a subsequent behavior, while multiple behaviors on different items may jointly reveal latent user interests. Adding behavior embeddings at the input can distinguish behavior types but cannot fully explain how behaviors influence one another. Relation modeling therefore aims to capture dependencies among interactions so that the importance of historical information can be adjusted according to behavior types and context.

Within Transformer architectures, such relations can be incorporated by extending attention computation. Let q_i and k_j denote the query and key representations of the i -th and j -th interactions. A simplified behavior-aware attention weight can be written as

$$\alpha_{ij} = \text{softmax} \left(\frac{q_i k_j^T}{\sqrt{d}} + r_{b_i, b_j} \right) \quad (3)$$

Where r_{b_i, b_j} represents relation information between behaviors b_i and b_j . Compared with standard self-attention, the association between two historical interactions is determined not only by content similarity but also by behavior types and their relations, strengthening the modeling of cross-behavior dependencies.

Early methods primarily relied on Transformer attention to learn behavior relations. MB-STR captures transition patterns among behaviors through behavior-specific sequence representations and multi-behavior pattern learning [13]. PBAT introduces behavior correlations into attention computation and combines them with item-level collaborative information and temporal factors, allowing historical behaviors to receive different weights according to the current user state [16]. These approaches mainly learn relations within an individual user sequence and make limited use of global behavior associations beyond that sequence.

Subsequent studies increasingly combined graph structures with Transformer to obtain richer relational information. MBHT uses a multi-scale Transformer to model local sequences and a hypergraph to capture high-order relations among behaviors and items, supplementing global information that is difficult to obtain from a single sequence [14]. TGT incorporates temporal information into graph structures and uses a Temporal Graph Transformer to model how cross-behavior relations evolve over time [15]. GPG4HSR constructs global and personalized graphs to learn behavior-transition patterns shared across users and user-specific behavioral characteristics, respectively [21].

As relation modeling became more fine-grained, the modeling unit expanded from behavior-level relations to specific interactions. M-GPT treats item-behavior interactions as graph nodes and uses graph information to enhance Transformer sequence representations [22]. HEML employs a multi-behavior hypergraph to capture high-order dependencies among interactions and combines them with multi-interest representations [23]. More recently, GXNSRec uses a Graph Cross Network to explicitly compute cross-features among heterogeneous behaviors and combines it with behavior-aware self-attention for long-range behavioral associations [24]. For long sequences and industrial settings, TGA selects key behavior transitions at the item, category,

and neighborhood levels and performs graph attention over a sparse transition graph, enabling multi-behavior transition modeling with linear complexity [25].

3.3. Temporal Modeling

Temporal information in multi-behavior sequences indicates not only the order of interactions but also the degree to which historical behaviors contribute to current prediction. The same behavior may have different predictive value depending on when it occurs, and different behavior types can exhibit different temporal effects. For example, a recent click may better reflect current interest, whereas an earlier favorite or purchase may represent a more stable long-term preference. Temporal modeling therefore needs to consider interaction order, time intervals, and differences in interest evolution across temporal ranges.

Early time-aware approaches explicitly introduced temporal information into self-attention. TiSASRec jointly models positional relations and interaction time intervals so that attention weights vary with temporal distance, laying a foundation for later time-aware Transformers [3]. In multi-behavior settings, PBAT jointly incorporates temporal factors and behavior relations into attention computation and adjusts the importance of historical interactions according to behavior type and occurrence time [16]. TGT integrates time with graph structure through a Temporal Graph Transformer, thereby modeling the evolution of cross-behavior relations rather than treating such dependencies as static [15].

As user histories grow, a single temporal distance is insufficient to describe both short-term interests and long-term preferences, making multi-scale modeling another important direction. MBHT extracts behavior patterns over different sequence ranges with a multi-scale Transformer, allowing local interactions and longer histories to be modeled separately [14]. HEML combines dual-scale sequence encoding with multi-interest learning to distinguish user states across temporal ranges [23]. M-GPT links fine-grained interaction changes with longer-term interests through multi-scale sequence representations and multi-granularity preferences [22]. The focus of these approaches extends beyond when a behavior occurred to how information from different temporal ranges contributes to current interest.

Recent studies further refine dynamic modeling beyond basic temporal representations. GEAR incorporates temporal context into item-behavior interaction representations and uses temporal bias to characterize changes in the contribution of historical information over time [18]. CAMBSRec improves attention from both positional encoding and frequency features, enabling the model to capture local ordering changes together with relatively stable interest patterns [26].

3.4. Interest Modeling

Multi-behavior sequences often reflect long-term interests, short-term needs, and immediate intentions simultaneously. Behaviors also differ in information strength and relevance to the target behavior. An important challenge is therefore to identify useful interests from complex historical interactions and construct target-aware representations that reflect the current user state. Existing work mainly focuses on multi-interest representation, personalization, target-behavior constraints, and noise suppression.

For multi-interest and personalized modeling, MAINT separates user state into preference and intent, represents

multiple concurrent interests from different aspects, and extracts current intent from behavior sequences [27]. PBAT focuses on differences in behavior patterns across users and learns user-specific behavior representations through a personalized behavior-pattern generator [16]. HEML combines multi-scale sequence modeling with multi-interest learning to extract interests at different temporal ranges and uses a multi-behavior hypergraph to supplement high-order behavior relations [23]. M-GPT adopts multi-granularity modeling to combine local interaction patterns with longer-term interests, improving the characterization of interest evolution [22].

Target-behavior information has also been increasingly used to constrain interest representations. NextIP jointly optimizes next-item prediction and purchase prediction, allowing the target behavior to participate in the aggregation of behavior-specific subsequences and producing user representations more closely related to final conversion intention [28]. GHTID addresses irrelevant interests introduced by auxiliary behaviors by extracting multi-behavior information from a global heterogeneous graph and using target-interest denoising to suppress components weakly related to the current prediction target [29]. END4Rec further focuses on redundant behaviors and noise in long sequences, decoupling effective information from noise to reduce the influence of irrelevant interactions while improving modeling efficiency [30]. These methods indicate that the value of multi-behavior information depends less on its quantity than on the ability to identify the portions truly relevant to the target behavior.

Interest modeling has also been extended with data augmentation, self-supervised learning, and generative methods. MBASR addresses multi-behavior data sparsity through several behavior-aware augmentation operations and a position-sampling strategy that controls perturbations to the original sequence, allowing the framework to be integrated with different MBSR models [31]. MBCCCL constructs collaborative contrastive signals from multi-behavior interactions and attribute information to enhance personalized sequence representations [32]. MBGen reformulates MBSR as a generative task: it first predicts the next behavior type and then generates the next item conditioned on the target behavior, with behaviors and items interleaved in a unified autoregressive sequence [33]. M³BSR further introduces multimodal information such as text and images, applies conditional diffusion to handle modality and behavior noise, and uses a mixture-of-experts mechanism to extract cross-behavior and cross-modal interests [34]. Transformer-based cross-layer contrastive learning constructs auxiliary supervision from representations at different network layers to strengthen information exchange among hierarchical behavior features [35]. FatsMB adopts latent diffusion to learn behavior-agnostic preferences in a unified latent space and transfer them toward target-behavior-specific preferences, while combining behavior-aware RoPE with multi-condition guided denoising for preference generation [36]. Overall, these studies extend MBSR from basic behavior utilization toward stronger interest representation and more selective use of historical information.

4. Conclusion And Future Directions

In recent years, Transformer has become a major modeling approach for MBSR. Compared with conventional sequential recommendation that focuses only on item order, recent

methods incorporate behavior types, temporal information, and dependencies among interactions so that heterogeneous user actions can be treated differently. The field has progressed from simple behavior embeddings to richer modeling of behavior relations, temporal dynamics, and user interests, while also incorporating graph structures, denoising, and contrastive learning. These studies demonstrate that multi-behavior information provides valuable evidence for user-interest estimation, but how to exploit such information effectively remains a central research issue.

Several challenges remain insufficiently addressed. First, many methods rely on fixed behavior-type representations, even though the same behavior may have different meanings for different users and contexts; for instance, a click can indicate explicit interest or merely an accidental action. Second, irrelevant information in auxiliary behaviors can interfere with target-behavior prediction, so using more behaviors does not necessarily yield better performance. Third, as user histories grow, Transformer must process longer interaction sequences, increasing both computational cost and exposure to irrelevant history. Finally, real recommendation platforms often need to support multiple target behaviors simultaneously, and these objectives may share useful information while also interfering with one another.

Future research may proceed in the following directions:

More accurate behavior representation: Beyond behavior type, dwell time, interaction intervals, item content, and contextual information can be incorporated to infer the intent expressed by each interaction more accurately.

Auxiliary-behavior selection and denoising: Historical behaviors should be evaluated according to the current recommendation target so that weakly related signals, such as accidental clicks or incidental browsing, exert less influence on prediction.

Efficient long-sequence modeling: Long-term interests should be retained while ineffective historical interactions are reduced through history selection, sequence compression, or efficient attention mechanisms.

Multi-target and content-aware modeling: Relations among different target behaviors can be modeled jointly with item content such as text and images to improve recommendation in complex scenarios.

In summary, Transformer has strengthened the ability of MBSR to model complex user histories, but performance still depends on correctly distinguishing behaviors, selecting useful historical evidence, and exploiting information according to the target task. Future work should consider not only recommendation accuracy but also computational efficiency, noise robustness, and applicability to real-world systems, enabling MBSR to be deployed more reliably in practical recommendation services.

Acknowledgements

This work is funded by the humanities and social sciences research project of the ministry of education (No.24YJA870003).

References

- [1] Kang, W. C., & McAuley, J. (2018). Self-attentive sequential recommendation. In 2018 IEEE International Conference on Data Mining (ICDM) (pp. 197–206).

- [2] Sun, F., Liu, J., Wu, J., et al. (2019). BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (pp. 1441–1450).
- [3] Li, J., Wang, Y., & McAuley, J. J. (2020). Time interval aware self-attention for sequential recommendation. In Proceedings of the 13th ACM International Conference on Web Search and Data Mining (pp. 322–330).
- [4] Zhou, K., Wang, H., Zhao, W. X., et al. (2020). S3-Rec: Self-supervised learning for sequential recommendation with mutual information maximization. In Proceedings of the 29th ACM International Conference on Information and Knowledge Management (pp. 1893–1902).
- [5] Xie, X., Sun, F., Liu, Z., et al. (2022). Contrastive learning for sequential recommendation. In 2022 IEEE 38th International Conference on Data Engineering (ICDE) (pp. 1259–1273).
- [6] Qiu, R., Huang, Z., Yin, H., et al. (2022). Contrastive learning for representation degeneration problem in sequential recommendation. In Proceedings of the 15th ACM International Conference on Web Search and Data Mining (pp. 813–823).
- [7] Chen, X., Li, Z., Pan, W., et al. (2026). A survey on multi-behavior sequential recommendation. *Science China Information Sciences*, 69(3), 131101.
- [8] Kim, K., Kim, S., Lee, G., et al. (2026). A comprehensive survey on multi-behavior recommender systems: Extended taxonomy and recent advances. *International Journal of Data Science and Analytics*, 22(1), 40.
- [9] Zhou, M., Ding, Z., Tang, J., et al. (2018). Micro behaviors: A new perspective in e-commerce recommender systems. In Proceedings of the 11th ACM International Conference on Web Search and Data Mining (pp. 727–735).
- [10] Li, Z., Zhao, H., Liu, Q., et al. (2018). Learning from history and present: Next-item recommendation via discriminatively exploiting user behaviors. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1734–1743).
- [11] He, M., Pan, W., & Ming, Z. (2022). BAR: Behavior-aware recommendation for sequential heterogeneous one-class collaborative filtering. *Information Sciences*, 608, 881–899.
- [12] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (pp. 5998–6008).
- [13] Yuan, E., Guo, W., He, Z., et al. (2022). Multi-behavior sequential transformer recommender. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1642–1652).
- [14] Yang, Y., Huang, C., Xia, L., et al. (2022). Multi-behavior hypergraph-enhanced transformer for sequential recommendation. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 2263–2274).
- [15] Xia, L., Huang, C., Xu, Y., et al. (2023). Multi-behavior sequential recommendation with temporal graph transformer. *IEEE Transactions on Knowledge and Data Engineering*, 35(6), 6099–6112.
- [16] Su, J., Chen, C., Lin, Z., et al. (2023). Personalized behavior-aware transformer for multi-behavior sequential recommendation. In Proceedings of the 31st ACM International Conference on Multimedia (pp. 6321–6331).
- [17] Elsayed, S., Rashed, A., & Schmidt-Thieme, L. (2024). Multi-behavioral sequential recommendation. In Proceedings of the 18th ACM Conference on Recommender Systems (pp. 902–906).
- [18] Jiang, J., Zhang, K., Kang, J., et al. (2025). GEAR: Generalized alternating regressor for multi-behavior sequential recommendation. In Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 2870–2874).
- [19] Li, Y., Cheng, M., Luo, Y., et al. (2026). BLADE: A behavior-level data augmentation framework with dual fusion modeling for multi-behavior sequential recommendation. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 40, No. 18, pp. 15189–15197).
- [20] Chen, X., Luo, Y., Xu, Z., Pan, W., & Ming, Z. (2026). Behavior-aware attribute-infused sequential encoders for next-item recommendation. *ACM Transactions on Information Systems*, 44(5), 107, 1–24.
- [21] Chen, W., He, M., Ni, Y., et al. (2022). Global and personalized graphs for heterogeneous sequential recommendation by learning behavior transitions and user intentions. In Proceedings of the 16th ACM Conference on Recommender Systems (pp. 268–277).
- [22] He, C., Liu, Y., Li, Q., et al. (2025). Multi-grained preference enhanced transformer for multi-behavior sequential recommendation. In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 872–883).
- [23] Li, Q., Ma, H., Jin, W., et al. (2024). Hypergraph-enhanced multi-interest learning for multi-behavior sequential recommendation. *Expert Systems with Applications*, 255, 124497.
- [24] Chen, R., Fan, J., Wu, M., Cheng, R., & Chai, M. (2026). GXNSRec: Multi-behavior sequential recommender based on graph cross networks. *Expert Systems with Applications*, 297, 129387.
- [25] Jin, H., Yang, G., Chan, Z., et al. (2026). Multi-behavior sequential modeling with transition-aware graph attention network for e-commerce recommendation. In Proceedings of the ACM Web Conference 2026 (pp. 8681–8684).
- [26] Zhuang, B., Lan, Y., & Zhang, M. (2025). CAMBSRec: A context-aware multi-behavior sequential recommendation model. *Informatics*, 12(3), 79.
- [27] Liu, H., Ding, J., Zhu, Y., et al. (2023). Modeling multi-aspect preferences and intents for multi-behavioral sequential recommendation. *Knowledge-Based Systems*, 280, 111013.
- [28] Luo, J., He, M., Lin, X., et al. (2022). Dual-task learning for multi-behavior sequential recommendation. In Proceedings of the 31st ACM International Conference on Information and Knowledge Management (pp. 1379–1388).
- [29] Li, X., Chen, H., Yu, J., et al. (2024). Global heterogeneous graph and target interest denoising for multi-behavior sequential recommendation. In Proceedings of the 17th ACM International Conference on Web Search and Data Mining (pp. 387–395).
- [30] Han, Y., Wang, H., Wang, K., et al. (2024). Efficient noise-decoupling for multi-behavior sequential recommendation. In Proceedings of the ACM Web Conference 2024 (pp. 3297–3306).
- [31] Xiao, Q., Xiao, J., Pan, W., et al. (2026). MBASR: A generic framework for multi-behavior data augmentation in sequential recommendation. *ACM Transactions on Recommender Systems*, 4(3), 36, 1–28.
- [32] Chen, Y., Cao, Q., Huang, X., et al. (2024). Multi-behavior collaborative contrastive learning for sequential recommendation. *Complex & Intelligent Systems*, 10(4), 5033–5048.
- [33] Liu, Z., Hou, Y., & McAuley, J. J. (2024). Multi-behavior generative recommendation. In Proceedings of the 33rd ACM

- International Conference on Information and Knowledge Management (pp. 1575–1585).
- [34] Cui, X., Lu, W., Tong, Y., et al. (2025). Multi-modal multi-behavior sequential recommendation with conditional diffusion-based feature denoising. In Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1593–1602).
- [35] Zhao, L., Tian, B., Wang, T., et al. (2025). Enhancing multi-behavior sequential recommendation via transformer-based cross-layer contrastive learning. In Web Information Systems Engineering – WISE 2025 (pp. 17–27).
- [36] Yang, R., Li, X., Sheng, J., et al. (2026). From agnostic to specific: Latent preference diffusion for multi-behavior sequential recommendation. In Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (Vol. 1, pp. 1775–1786).