ISSN: 3126-2953
E-ISSN: 3126-3097

NEUMF–LSTM: IMPROVING TARGET BEHAVIOR PREDICTION IN MULTI-BEHAVIOR RECOMMENDER SYSTEMS

LE THI HANH HIEN[1,a], TRAN THU MAI[1],
NGUYEN THAI NGHE[2], TRAN THI TO LOAN[1]
[1]Vinh Long University of Technology Education, Vinh Long
[2]College of Information and Communication Technology-Can Tho University, Can Tho
[a]Corresponding author: hienlth@vlute.edu.vn

Reviewed: 04/11/2025; Accepted: 26/12/2025

ABSTRACT

This study proposes a NeuMF–LSTM model for multi-behavior recommender systems in e-commerce, integrating the nonlinear interaction learning capability of NeuMF with the sequential modeling strength of LSTM. Experiments on the Tianchi and Tmall datasets show that the proposed model improves purchase prediction accuracy compared to baseline methods, confirming its effectiveness.

Keywords: *recommender system, multi-behavior, e-commerce, NeuMF, LSTM*

1. INTRODUCTION

Modern recommendation systems have become essential tools for helping users find suitable products in the digital age. Previous studies often focus on predicting purchase behavior based on a single type of interaction. In reality, however, consumers engage in a sequence of actions, such as viewing (pv), favoriting (fav), adding items to the cart (cart), and making purchases (buy). This sequence provides valuable signals that can significantly improve the accuracy of purchase predictions. Deep learning models such as MLP, NeuMF, Wide & Deep, DeepFM, LSTM, and Transformer have demonstrated strong performance in recommendation systems. However, each individual model comes with its own limitations. To address this, the present study proposes a hybrid model that combines the nonlinear interaction learning capability of NeuMF with the sequential modeling strength of LSTM. This model, named RecNeuLSTM, is designed to capture both static and dynamic features from multi-behavior data.

Experimental results on two large-scale datasets, Tianchi and Tmall, show that the proposed model significantly improves the accuracy of purchase behavior prediction compared to baseline models.

2. CONTENT

2.1. Related Work

Recommendation systems are a major research direction in e-commerce, originally rooted in collaborative filtering and content-based filtering. However, these traditional approaches suffer from limitations such as data sparsity and poor generalization. With the rise of deep learning, various neural network models have been developed to address these issues. NeuMF [4] combines Generalized Matrix Factorization (GMF) and Multilayer Perceptron (MLP) to learn nonlinear user–item interactions. Wide & Deep [2] and DeepFM [3] further enhance modeling capability by integrating linear and nonlinear components, proving especially effective in learning high-dimensional feature representations. Despite their strengths, these models typically focus on a single type of behavior, usually purchases, and fail to fully leverage information from multiple user actions.

In sequential recommendation, LSTM [1] has been shown to effectively capture temporal dependencies, while Transformer [5] excels at modeling long-range relationships through self-attention mechanisms. However, LSTM struggles with complex nonlinear interactions, and Transformer models often require large-scale training data. Recent studies such as Disen-CGCN [10] and CAMBSRec [14] highlight the importance of jointly modeling both sequential and cross-behavior relationships.

A prominent trend is multi-behavior recommendation, where signals from various actions such as viewing, favoriting, adding to cart, and purchasing are integrated to improve prediction performance. Methods like MBGen [12], TOAR [13], Multi-Behavior Aware Recommendation [15], Hierarchical Fine-grained Multi-behavior Recommendation [16], and Transformer-based Collaborative Filtering [17] demonstrate the effectiveness of this approach. At the same time, they reveal challenges such as negative transfer, where auxiliary behaviors may introduce noise and degrade prediction quality. A recent survey [18] also emphasizes the need for hybrid models that can leverage the strengths of different approaches.

Overall, NeuMF is effective in modeling nonlinear interactions but lacks the ability to capture sequential patterns, while LSTM handles sequential data well but is limited in modeling complex interactions. Therefore, this study proposes a hybrid model, RecNeuLSTM, for multi-behavior recommendation systems, aiming to improve the accuracy of purchase prediction, the most critical objective in e-commerce.

2.2. Proposed Method

In e-commerce, user behaviors such as viewing (pv), favoriting (fav), adding to cart (cart), and purchasing (buy) typically occur in a logical sequence. Among these, purchase is the most important but also the least frequent behavior. Earlier actions, while less critical, provide valuable auxiliary signals that help improve prediction accuracy. Therefore, an effective recommendation system should model both the sequential relationships and cross-behavior interactions, while balancing static signals (nonlinear user–item interactions) and dynamic signals (temporal behavior sequences) to comprehensively capture user preferences and intentions.

The proposed RecNeuLSTM model is designed to leverage both static and dynamic features from multi-behavior data. Within the NeuMF framework, two parallel components are employed. The GMF branch captures linear relationships through element-wise multiplication of user and item embeddings, while the MLP branch learns nonlinear interactions by concatenating embeddings and passing them through fully connected layers. These two components produce a stable representation of user–item interactions. In parallel, the LSTM branch processes sequences of user behaviors to learn temporal dependencies among actions. The final hidden state of the LSTM reflects the dynamic relationships between behaviors, enabling the model to better capture the user's decision-making process.

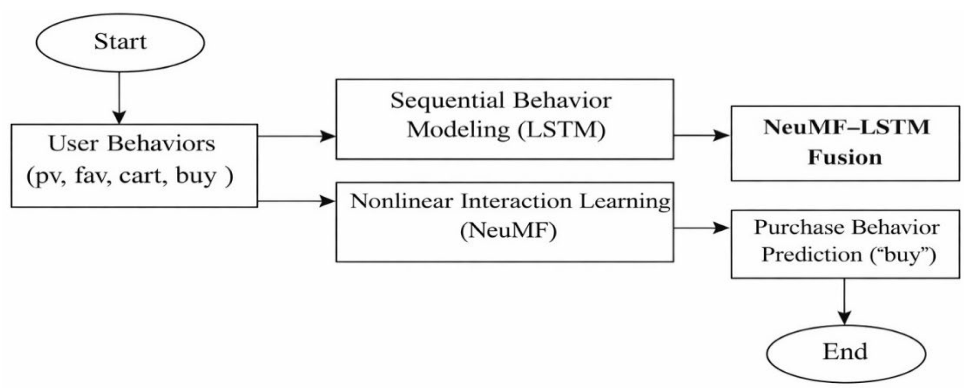

Figure 1. The workflow of the proposed RecNeuLSTM method for multi-behavior recommendation, from input behaviors (pv, fav, cart, buy) to purchase prediction.

At the fusion stage, the representations from GMF, MLP, and LSTM are concatenated into a unified feature vector. This combined representation is then passed through a fully connected layer followed by a sigmoid activation function to estimate the probability of each behavior. The NeuMF and LSTM components are trained jointly within a single framework, using a multi-task loss function to optimize both auxiliary behaviors and the primary target behavior simultaneously. The loss function is based on Binary Cross-Entropy (BCE), with a balancing weight αb applied to mitigate the imbalance among behaviors with different frequencies.

The RecNeuLSTM approach integrates two complementary sources of information within the recommendation system. It not only captures complex user–item interactions but also models the temporal evolution of user behavior. These two branches are combined through a fusion layer, resulting in a more comprehensive representation of user preferences and behavioral patterns, thereby improving prediction accuracy.

Nhánh NeuMF:

GMF: $z_{GMF} = u \odot i$

MLP: $x_{MLP} = [u;i] \Rightarrow z_{MLP} = MLP(x_{MLP})$

$$z_{NeuMF} = [z_{GMF} ; z_{MLP}]$$

Where:

$\odot$: Hadamard product (element-wise multiplication)

z_{GMF}: representation capturing linear interaction patterns between user and item

z_{MLP}: representation capturing complex nonlinear interactions

z_{NeuMF}: concatenated vector from GMF and MLP

LSTM Branch

$$H_t, c_t = LSTM(x_t , h_t-1 , c_t-1)$$

Where:

x_t: input vector at time step t

h_t: hidden state at time step t

c_t: cell state at time step t.

Fusion NeuMF + LSTM:

$$z_{fusion} = [z_{NeuMF} ; h_t]$$

Trong đó:

z_{fusion}: vector tổng hợp cả tương tác phi tuyến tĩnh và biểu diễn tuần tự.

Biểu diễn tổng hợp z_{fusion} được đưa qua lớp fully-connected cuối cùng để ước lượng xác suất của hành vi mục tiêu:

$$\hat{y} = \sigma(W_f . z_{fusion} + b_f)$$

Where:

W_f : weight matrix of the final fully connected layer

b_f : bias term

σ: sigmoid activation function mapping outputs to [0,1]

ŷ: predicted probability of the target behavior

To evaluate the effectiveness of the model, this study adopts three widely used metrics: Precision (P), Recall (R), and NDCG (Normalized Discounted Cumulative Gain). Precision measures the accuracy of the recommended items, while Recall reflects the model's ability to retrieve relevant items. NDCG evaluates the ranking quality by assigning higher importance to correctly predicted items that appear at top positions. Together, these metrics provide a comprehensive assessment of the recommendation system's performance.

Chỉ số	Công thức
Precision (P)	$P = \frac{\lvert Recommended \cap Relevant \rvert}{\lvert Recommended \rvert}$
Recall (R)	$R = \frac{\lvert Recommended \cap Relevant \rvert}{\lvert Relevant \rvert}$
NDCG@K	$NDCG@K = \frac{DCG@K}{IDCG@K}$

2.3. Experiments

2.3.1. Datasets

The experiments were conducted on two widely used e-commerce datasets:

Table 1. Description of Experimental Datasets

Dataset	Item	User	Interactions	Behaviors (pv, fav, cart, buy)
Tianchi	500,900	25,000	4,619,389	✓
Tmall	99,037	147,894	7,658,926	✓

Both datasets share a common characteristic: an imbalance among behavior types. Specifically, the "view" behavior dominates, while the more important "purchase" behavior appears much less frequently. This imbalance poses a challenge in effectively leveraging auxiliary behaviors to support the prediction of the target behavior. The data is split into 70% for training, 10% for validation, and 20% for testing. The model is trained using the Adam optimizer with a learning rate of 0.001, a batch size of 256, and a dropout rate of 0.3 to prevent overfitting. The number of training epochs is set to 5. To address the imbalance among behaviors, weighted negative sampling is applied, and the coefficient αb in the loss function is adjusted to assign greater importance to the "purchase" behavior.

The model is trained for 5 epochs. The observed loss and accuracy indicate that the model converges quickly; therefore, limiting training to 5 epochs helps

ISSN: 3126-2953
E-ISSN: 3126-3097

prevent overfitting while also reducing computational cost.

2.3.2. Experimental Results

The model's performance during training is evaluated using the metrics Precision (P), Recall (R), and NDCG. The results for the proposed RecNeuLSTM model are presented as follows:

Table 2. Experimental results on the Tianchi dataset

Behavior	Model	@5			@10		
		P	R	NDCG	P	R	NDCG
pv	MLP	0.9791	0.1149	0.9821	0.9699	0.2197	0.9793
	NeuMF	0.9829	0.1163	0.9856	0.9738	0.2219	0.9831
	Wide&Deep	0.9743	0.1133	0.9772	0.9652	0.2178	0.9743
	DeepFM	0.9768	0.1141	0.9795	0.9676	0.2190	0.9767
	LSTM	0.9804	0.1150	0.9833	0.9713	0.2203	0.9805
	Transformer	0.9745	0.1130	0.9775	0.9653	0.2174	0.9745
	RecNeuLSTM	0.9865	0.1177	0.9892	0.9774	0.2239	0.9869
fav	MLP	0.2705	0.1859	0.3209	0.2446	0.3109	0.3433
	NeuMF	0.3193	0.2241	0.3848	0.2811	0.3558	0.4030
	Wide&Deep	0.2618	0.1791	0.3112	0.2360	0.2968	0.3315
	DeepFM	0.3059	0.2112	0.3658	0.2685	0.3391	0.3831
	LSTM	0.3136	0.2227	0.3784	0.2780	0.3530	0.3985
	Transformer	0.2646	0.1914	0.3234	0.2356	0.3090	0.3429
	RecNeuLSTM	0.3788	0.2729	0.4608	0.3278	0.4163	0.4778
cart	MLP	0.0890	0.2090	0.1597	0.0813	0.3653	0.2173
	NeuMF	0.1274	0.2893	0.2387	0.1022	0.4541	0.2976
	Wide&Deep	0.0895	0.1974	0.1592	0.0755	0.3315	0.2073
	DeepFM	0.1124	0.2505	0.2065	0.0900	0.3925	0.2578
	LSTM	0.1025	0.2257	0.1813	0.0859	0.3742	0.2348
	Transformer	0.0639	0.1477	0.1098	0.0611	0.2817	0.1589
	RecNeuLSTM	0.1406	0.3158	0.2620	0.1083	0.4739	0.3180
buy	MLP	0.4483	0.2609	0.4681	0.3801	0.4320	0.4823

ISSN: 3126-2953
E-ISSN: 3126-3097

Behavior	Model	@5			@10		
		P	R	NDCG	P	R	NDCG
	NeuMF	0.4760	0.2776	0.4994	0.3978	0.4525	0.5097
	Wide&Deep	0.4267	0.2483	0.4461	0.3627	0.4133	0.4604
	DeepFM	0.4646	0.2706	0.4884	0.3867	0.4398	0.4968
	LSTM	0.4485	0.2604	0.4697	0.3794	0.4313	0.4828
	Transformer	0.4056	0.2361	0.4238	0.3474	0.3971	0.4400
	RecNeuLSTM	0.5688	0.3312	0.5982	0.4630	0.5213	0.5988

The results on the Tianchi dataset show that the NeuMF–LSTM model outperforms individual models across all four behaviors (pv, fav, cart, buy). In particular, for the purchase behavior, the model achieves P@5 = 0.5688, R@10 = 0.5213, and NDCG@10 = 0.5988, which are significantly higher than those of NeuMF (P@5 = 0.4760, R@10 = 0.4525, NDCG@10 = 0.5097). This improvement indicates that the combination of NeuMF and LSTM enables the model to effectively capture both nonlinear user–item interactions and sequential behavior patterns, thereby enhancing prediction accuracy. Overall, NeuMF–LSTM demonstrates strong performance and stability, highlighting the clear advantages of the hybrid approach in user behavior prediction. To further validate its robustness, the study extends the evaluation to the larger and more diverse Tmall dataset.

Table 3. Experimental results on the Tmall dataset

Behavior	Model	@5			@10		
		P	R	NDCG	P	R	NDCG
pv	MLP	0.9567	0.1821	0.9605	0.9399	0.3487	0.9547
	NeuMF	0.9673	0.1853	0.9708	0.9500	0.3534	0.9650
	Wide&Deep	0.9517	0.1808	0.9552	0.9358	0.3468	0.9501
	DeepFM	0.9624	0.1837	0.9658	0.9458	0.3514	0.9604
	LSTM	0.9540	0.1812	0.9580	0.9373	0.3474	0.9522
	Transformer	0.9487	0.1797	0.9529	0.9320	0.3449	0.9469
	RecNeuLSTM	0.9712	0.1863	0.9745	0.9536	0.3550	0.9686
fav	MLP	0.1632	0.2139	0.2222	0.1474	0.3764	0.2738
	NeuMF	0.1933	0.2604	0.2707	0.1672	0.4296	0.3224

Behavior	Model	@5			@10		
		P	R	NDCG	P	R	NDCG
	WideandDeep	0.1795	0.2439	0.2516	0.1568	0.4070	0.3024
	DeepFM	0.1982	0.2696	0.2812	0.1695	0.4397	0.3328
	LSTM	0.1728	0.2285	0.2373	0.1538	0.3934	0.2892
	Transformer	0.1379	0.1851	0.1881	0.1295	0.3378	0.2394
	RecNeuLSTM	0.2310	0.3147	0.3318	0.1907	0.4897	0.3820
cart	MLP	0.2128	0.2443	0.2705	0.1911	0.4261	0.3290
	NeuMF	0.2726	0.3159	0.3567	0.2278	0.5028	0.4103
	WideandDeep	0.2195	0.2522	0.2807	0.1941	0.4312	0.3371
	DeepFM	0.2634	0.3033	0.3446	0.2215	0.4864	0.3974
	LSTM	0.2081	0.2376	0.2631	0.1879	0.4187	0.3217
	Transformer	0.1820	0.2086	0.2267	0.1729	0.3903	0.2891
	RecNeuLSTM	0.2843	0.3302	0.3752	0.2355	0.5176	0.4278
buy	MLP	0.3170	0.3917	0.4082	0.2463	0.5976	0.4949
	NeuMF	0.3461	0.4289	0.4507	0.2606	0.6320	0.5352
	WideandDeep	0.3098	0.3829	0.3979	0.2428	0.5893	0.4852
	DeepFM	0.3399	0.4206	0.4425	0.2570	0.6233	0.5268
	LSTM	0.3077	0.3803	0.3938	0.2424	0.5886	0.4821
	Transformer	0.2974	0.3674	0.3786	0.2364	0.5744	0.4668
	RecNeuLSTM	0.3634	0.4505	0.4766	0.2691	0.6521	0.5596

The results on the Tmall dataset further confirm the effectiveness of the NeuMF–LSTM model compared to baseline approaches. For the purchase behavior (buy), the hybrid model achieves P@5 = 0.3634, R@10 = 0.6521, and NDCG@10 = 0.5596. Intermediate behaviors such as favoriting (fav) and adding to cart (cart) also show consistent improvements, particularly in terms of Recall and NDCG, indicating a stronger ability to learn from sequential user behavior. The integration of NeuMF and LSTM enables the model to capture both static features and temporal dynamics, thereby improving prediction accuracy and ranking performance on large-scale, noisy, and multi-behavior data. Compared to the results on the Tianchi dataset, the RecNeuLSTM model continues to demonstrate stable performance and

ISSN: 3126-2953
E-ISSN: 3126-3097

better generalization when applied to more complex data.

In addition, the decreasing loss and steadily increasing accuracy during both training and validation indicate that the model is effectively optimized, with no significant signs of overfitting. This further supports the robustness of the proposed hybrid architecture.

Tianchi	Tmall

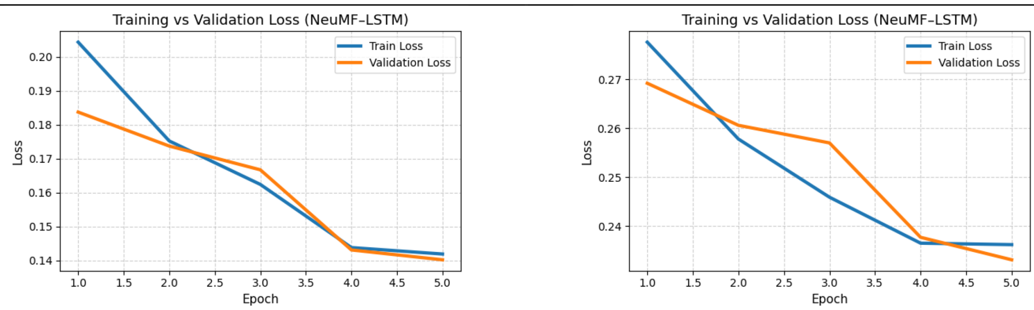

Figure 3. Training and validation loss curves over 5 epochs.

Figure 3 shows that the loss value gradually decreases across epochs, indicating that the model converges well and achieves stable optimization during training.

Tianchi	Tmall

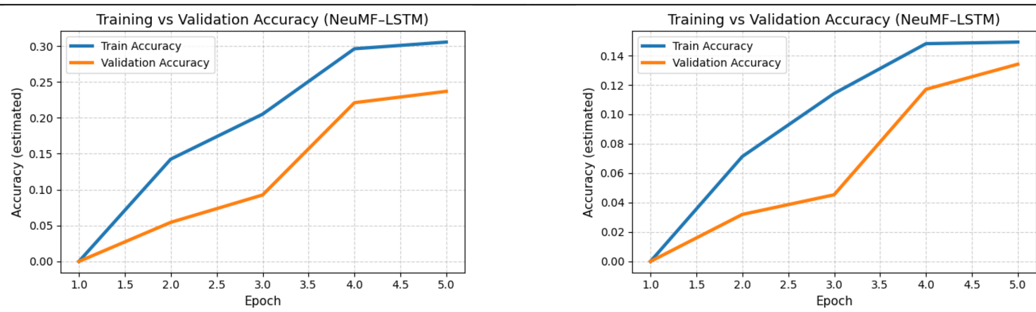

Figure 4. Training and validation accuracy curves over 5 epochs.

Figure 4 shows that the accuracy steadily increases and stabilizes over epochs, reflecting the model's effective learning capability without exhibiting signs of overfitting.

3. CONCLUSION

This study conducts experiments on two large-scale e-commerce datasets (Tianchi and Tmall) to evaluate six prominent deep learning models and introduces a hybrid NeuMF–LSTM approach for multi-behavior recommendation. The results demonstrate that the proposed method effectively leverages the strengths of NeuMF in modeling nonlinear interactions and LSTM in capturing sequential behavior patterns, thereby improving the accuracy of purchase prediction, which is the primary objective in e-commerce systems.

In future work, we plan to focus on more effective integration of multiple behaviors and address the issue of negative transfer. This will enable better utilization of auxiliary behavioral signals while reducing noise, ultimately enhancing both the accuracy and stability of multi-behavior recommendation systems.

REFERENCE

[1] S. Hochreiter and J. Schmidhuber. "Long Short-Term Memory." in Neural Computation. vol. 9. no. 8. pp. 1735-1780. 15 Nov. 1997. doi: 10.1162/neco.1997.9.8.1735.

[2] Cheng. Heng-Tze. et al. "Wide & deep learning for recommender systems." *Proceedings of the 1st workshop on deep learning for recommender systems*. 2016

[3] Guo. Huifeng. et al. "DeepFM: a factorization-machine based neural network for CTR prediction." *arXiv preprint arXiv:1703.04247* (2017).

[4] He. Xiangnan. et al. "Neural collaborative filtering." *Proceedings of the 26th international conference on world wide web*. 2017.

[5] Vaswani. Ashish. et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

[6] Xia. Lianghao. et al. "Knowledge-enhanced hierarchical graph transformer network for multi-behavior recommendation." *Proceedings of the AAAI conference on artificial intelligence*. Vol. 35. No. 5. 2021.

[7] Xu. Jingcao. et al. "Multi-behavior self-supervised learning for recommendation." *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval*. 2023.

[8] Liu. Zihan. Yupeng Hou. and Julian McAuley. "Multi-behavior generative recommendation." *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 2024.

[9] Yan. Mingshi. et al. "Behavior-contextualized item preference modeling for multi-behavior recommendation." *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval*. 2024.

[10] Cheng. Zhiyong. et al. "Disentangled cascaded graph convolution networks for multi-behavior recommendation." *ACM Transactions on Recommender Systems* 2.4 (2024): 1-27.

[11] F. Li. Y. Liu. L. Xu. Z. Qiao and Y. Li. "Convolutional Concatenation Fusion Imaging Method for ERT/UTT Dual-Modality Tomography." in IEEE Access. vol. 12. pp. 118099-118108. 2024. doi: 10.1109/ACCESS.2024.3447477.

[12] Liu. Zihan. Yupeng Hou. and Julian McAuley. "Multi-behavior generative recommendation." *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 2024.

[13] Yuan. Y.. Zhou. Y.. Chen. X.. Xiong. Q.. & Okere. H. C. "Enhancing Recommendation Diversity and Novelty with Bi-LSTM and Mean Shift Clustering. ". *Electronics*. 2024.

[14] Kim. Doyeon et al. "Accurate multi-behavior sequence-aware recommendation via graph convolution networks." *PloS one* vol. 20.1 e0314282. 7 Jan. 2025. doi:10.1371/journal.pone.0314282.

[15] Du. Y.. Yu. Z. & Wang. H. Multi-behavior aware recommendation with joint contrastive learning and reinforced negative sampling. *Complex Intell. Syst.* **11**. 352 (2025). https://doi.org/10.1007/s40747-025-01970-1.

[16] Zhu. Kaiyao et al. "Hierarchical fine-grained multi-behavior recommendation with behavior-aware contrastive learning." *Neural networks : the official journal of the International Neural Network Society*. vol. 192 107912. 28 Jul. 2025. doi:10.1016/j.neunet.2025.107912

[17] Khan. Hikmat Ullah et al. "A transformer-based architecture for collaborative filtering modeling in personalized recommender systems." Scientific reports vol. 15.1 24503. 8 Jul. 2025. doi:10.1038/s41598-025-08931-1.

[18] Kim. Kyungho. et al. "Multi-Behavior Recommender Systems: A Survey." *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Singapore: Springer Nature Singapore. 2025.