

FEDIN: Frequency-Enhanced Deep Interest Network for Click-Through Rate Prediction

Zenan Dai*
Tsinghua University
Shenzhen, China
dzn24@mails.tsinghua.edu.cn

Jinpeng Wang*[†]
Harbin Institute of Technology,
Shenzhen
Shenzhen, China
wangjp26@gmail.com

Junwei Pan
Tencent
Shenzhen, China
jonaspan@tencent.com

Dapeng Liu
Tencent
Shenzhen, China
rocliu@tencent.com

Lei Xiao
Tencent
Shenzhen, China
shawnxiao@tencent.com

Shu-Tao Xia
Tsinghua University
Shenzhen, China
xiast@sz.tsinghua.edu.cn

Abstract

Sequential recommendation models often struggle to capture latent periodic patterns in user interests, primarily due to the noise inherent in time-domain behavioral data. While frequency-domain analysis offers a global perspective to address this, existing approaches typically treat user sequences in isolation, overlooking the crucial context of the target item. In this work, we present a novel empirical observation: user attention scores exhibit distinct spectral entropy distributions when conditioned on positive versus negative target items. Specifically, true user interests manifest as highly concentrated spectral patterns with lower entropy in the frequency domain, whereas irrelevant behaviors appear as high-entropy noise. Leveraging this insight, we propose the Frequency-Enhanced Deep Interest Network (FEDIN). FEDIN introduces a frequency-domain branch that utilizes a target-aware spectrum filtering mechanism to isolate these periodic interest signals. Extensive experiments on three public datasets demonstrate that FEDIN consistently outperforms state-of-the-art sequential recommendation baselines, demonstrating superior robustness against noise. We have released our code at: <https://github.com/otokoneko/FEDIN>.

CCS Concepts

• Information systems → Recommender systems.

Keywords

Target Attention, Sequential Recommendation, CTR Prediction, Frequency Domain

ACM Reference Format:

Zenan Dai, Jinpeng Wang, Junwei Pan, Dapeng Liu, Lei Xiao, and Shu-Tao Xia. 2026. FEDIN: Frequency-Enhanced Deep Interest Network for Click-Through Rate Prediction. In *Proceedings of the 49th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '26)*, July 20–24, 2026, Melbourne, VIC, Australia. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3805712.3809861>

*Equal Contribution.

[†]Corresponding author.



This work is licensed under a Creative Commons Attribution 4.0 International License. *SIGIR '26, Melbourne, VIC, Australia*

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2599-9/2026/07

<https://doi.org/10.1145/3805712.3809861>

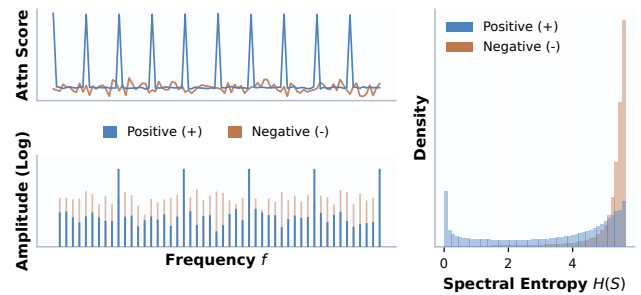


Figure 1: Empirical observation of target-conditioned spectral patterns. (Left) A case study comparing attention scores under different targets: positive items tend to elicit more pronounced periodic resonance in the user sequence, manifesting as concentrated spectral peaks. (Right) Statistical distribution of spectral entropy $H(S) = -\sum p_k \log_2 p_k$ across datasets, showing that positive samples exhibit generally lower entropy signatures compared to negative ones.

[†]26), July 20–24, 2026, Melbourne, VIC, Australia. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3805712.3809861>

1 Introduction

Click-Through Rate (CTR) prediction is a core component of modern recommendation systems, serving as a critical bridge between user preferences and candidate items. The efficacy of CTR models largely depends on how well they extract user interests from historical behavior sequences [21]. While deep sequential models, including RNNs [5, 18, 22], Transformers and SSMs [1, 6, 9, 13, 20], have achieved remarkable success by capturing temporal dependencies, they face inherent challenges in handling real-world user behaviors. User history is often a mixture of long-term periodic habits and short-term stochastic noise. Purely time-domain approaches, including those with target attention mechanisms [4, 21–23], tend to be sensitive to such point-wise noise and may struggle to decouple stable periodic patterns from random clicks in long sequences.

To tackle the limitations of time-domain modeling, frequency-domain analysis becomes a promising direction, showing success in long-term time series forecasting [2, 17, 25, 26]. The Fast Fourier

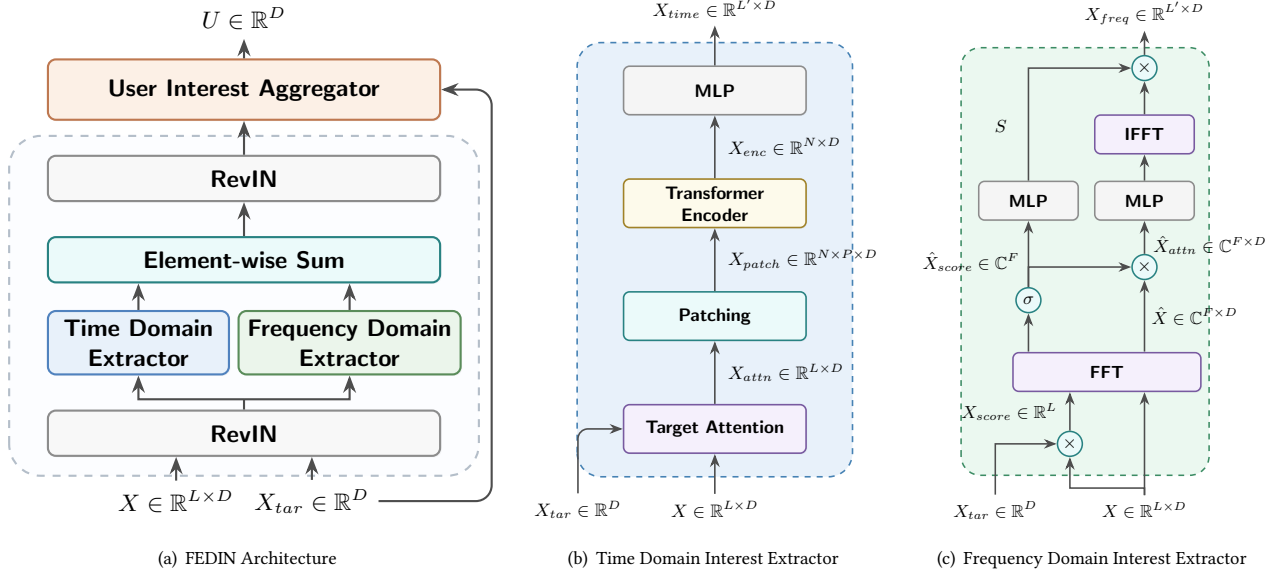


Figure 2: The architecture of FEDIN.

Transform offers a global view of signal periodicity, making it naturally suitable for denoising and pattern recognition. Inspired by this, recent works have explored frequency-domain analysis to address noise and sparsity. Early attempts like **FMLP-Rec** [24] and **FEARec** [3] introduced learnable filters to suppress stochastic noise. More recently, **DIFF** [7] has emerged as a state-of-the-art representative, utilizing discrete Fourier transforms to prioritize long-term stable interests through multi-sequence filtering. However, even advanced models like DIFF apply frequency analysis to the user sequence independently of the candidate item. We argue that this target-agnostic filtering is suboptimal because a user’s behavior sequence is polymorphic. Without the context of a specific target item, the model cannot effectively distinguish between the frequency components representing the true signal and those representing background noise.

In this paper, we revisit frequency-domain recommendation from a target-aware perspective. Through a statistical analysis of attention scores derived from a standard target attention backbone, we present a key empirical observation that the Target Attention mechanism exhibits distinct characteristics in the frequency domain. When these scores are projected into the frequency domain, true user interests manifest as highly concentrated spectral patterns with lower entropy, whereas irrelevant behaviors exhibit high-entropy distributions resembling white noise. As shown in Fig. 1, when we project the attention scores of user behaviors conditioned on different items into the frequency domain, a clear pattern emerges. Specifically, true user interests manifest as highly concentrated spectral patterns with lower entropy, whereas irrelevant behaviors exhibit high-entropy distributions resembling white noise. This insight suggests that the target item acts as a natural frequency selector. Simply calculating the spectrum of the raw behavior sequence overlooks this critical conditional phenomenon.

Building on this discovery, we propose the Frequency-Enhanced Deep Interest Network (**FEDIN**). Crucially, we recognize that frequency analysis should complement, not replace, time-domain modeling. While frequency features capture global periodicities and resist noise, time-domain features excel at capturing delicate local sequential evolution. Therefore, FEDIN adopts a dual-branch architecture that effectively fuses these two perspectives. The frequency branch utilizes a learnable spectrum filtering mechanism to isolate the structured harmonic patterns identified in our analysis, while the time branch maintains the granularity of sequential interactions. This design allows FEDIN to leverage the robustness of the frequency domain without sacrificing the detail of the time domain. Our contributions are summarized as follows:

- We empirically identify that user interests exhibit distinct low-entropy signatures in the frequency domain when conditioned on target items, providing a new theoretical grounding for spectral recommendation.
- We propose FEDIN, a hybrid framework that integrates a target-aware frequency branch with traditional time-domain modeling, effectively combining global noise robustness with local sequential precision.
- Experiments on three public datasets demonstrate that FEDIN consistently outperforms state-of-the-art sequential baselines, demonstrating superior effectiveness and robustness.

2 Method

2.1 Overall Framework

The overall architecture of FEDIN is illustrated in Fig. 2. Following the Embedding&MLP paradigm, FEDIN takes the user behavior sequence $X \in \mathbb{R}^{L \times D}$ and the target item $X_{tar} \in \mathbb{R}^D$ as inputs to predict the user interest representation $U \in \mathbb{R}^D$. Recognizing that

user behaviors contain both evolving sequential dependencies and stable periodic patterns, FEDIN employs a dual-branch denoising architecture. Notably, this decoupled design allows the two branches to be executed in a highly parallelized manner, ensuring efficiency for real-time deployment. First, we apply Reversible Instance Normalization (RevIN) [8] as a fundamental preprocessing component to mitigate distribution shifts and the non-stationary nature of user behaviors. The Time-Domain Branch focuses on capturing the local sequential evolution of interests, while the Frequency-Domain Branch focuses on isolating global periodic resonance patterns via spectral filtering. Finally, the User Interest Aggregator dynamically fuses these views to derive the final representation.

2.2 Time-Domain Branch

This branch aims to extract interest evolution from the time-domain sequence. A naive application of Transformers on raw sequences is often inefficient and does not necessarily guarantee performance gains due to noise [19]. Therefore, we design a coarse-to-fine extraction process.

Coarse Filtering via Target Attention. To mitigate pointwise noise, we first apply a target-aware coarse filter:

$$\mathbf{X}_{\text{attn}} = \text{Softmax}(\mathbf{X}_{\text{tar}}\mathbf{X}^T/\alpha)^T \odot \mathbf{X}, \quad (1)$$

where \odot denotes element-wise product. Unlike conventional target attention that performs weighted pooling, we compute a weighted sequence $\mathbf{X}_{\text{attn}} \in \mathbb{R}^{L \times D}$ to preserve temporal resolution for subsequent fine-grained evolution modeling.

Fine-grained Evolution via Patching Transformer. While TA selects relevant items, it does not model the temporal dependencies among them. To capture how these specific interests evolve over time, we employ a Transformer encoder [16]. Crucially, to handle long sequences efficiently and capture local temporal semantics, we adopt a patching strategy [11]. Specifically, \mathbf{X}_{attn} is segmented into non-overlapping patches $\mathbf{X}_{\text{patch}} \in \mathbb{R}^{N \times P \times D}$. We pad the sequence with zeros if L is not divisible by P , then flatten and linearly project each patch to dimension D . A Transformer encoder then models the dependencies between these patches, yielding the time-domain interest representation \mathbf{X}_{time} . This design ensures that we not only identify what the user likes but also how that interest changes over time.

2.3 Frequency-Domain Branch

As discussed in Section 1, true user interests manifest as concentrated harmonic patterns in the frequency domain. This branch aims to recover these signals from noisy behaviors using a target-aware spectral filtering mechanism.

First, we calculate the raw target attention scores $\mathbf{X}_{\text{score}} \in \mathbb{R}^L$ to capture the relevance of each historical behavior to the target item:

$$\mathbf{X}_{\text{score}} = \mathbf{X}\mathbf{X}_{\text{tar}}^T/\alpha, \quad (2)$$

where α is the scaling factor. Let $\hat{\mathbf{X}}_{\text{score}} = \mathcal{F}(\mathbf{X}_{\text{score}}) \in \mathbb{C}^L$ represent the target-conditioned spectrum. Unlike prior works that transform the raw sequence \mathbf{X} , we transform these attention scores to ensure the spectrum is conditioned on the target context. The target-aware spectrum is computed as:

$$\hat{\mathbf{X}}_{\text{attn}} = \text{Softmax}(\text{Amp}(\hat{\mathbf{X}}_{\text{score}})) \otimes \mathcal{F}(\mathbf{X}), \quad (3)$$

where $\mathcal{F}(\cdot)$ is the FFT operation and \otimes denotes the broadcast-ing element-wise product along the dimension D . This operation highlights frequency components where the user shows strong resonance with the target item.

Learnable Spectral Filter. To isolate the signal from noise, we need a filter to suppress high-entropy frequencies. Instead of using fixed filters, we employ a Complex-Valued MLP as a learnable spectral filter, adhering to the standard complex formulation defined in [15] to strictly preserve the intrinsic phase-amplitude coupling. The MLP acts on the complex spectrum, adaptively amplifying dominant harmonic frequencies while suppressing noise:

$$\hat{\mathbf{X}}_{\text{freq}} = \mathcal{F}^{-1}(\text{MLP}(\hat{\mathbf{X}}_{\text{attn}})). \quad (4)$$

Here, the MLP learns the non-linear mapping from the noisy input spectrum to the clean interest spectrum.

Adaptive Resonance Scaling. Not all user-item pairs exhibit strong periodicity. Based on our empirical observation, strong interests show lower spectral entropy. To leverage this, we introduce an adaptive gating mechanism that scales the frequency branch output based on the spectral clarity:

$$\mathbf{X}_{\text{freq}} = \text{Sigmoid}(\text{MLP}(\text{Amp}(\hat{\mathbf{X}}_{\text{score}}))) \cdot \hat{\mathbf{X}}_{\text{freq}}. \quad (5)$$

This ensures the model relies more on the frequency view when clear resonance patterns are detected, and falls back to the time domain otherwise.

Complexity Analysis. The computational cost of the Frequency-Domain Branch is dominated by FFT operations with a complexity of $O(L \log L)$, where L is the sequence length. This provides a significant efficiency advantage over the $O(L^2)$ complexity of standard self-attention mechanisms, particularly as L increases.

2.4 User Interest Aggregator

The two branches provide complementary views, where \mathbf{X}_{time} captures local evolution and \mathbf{X}_{freq} captures global periodicity. We sum them to obtain the mixed representation \mathbf{X}_{mix} . To derive the final user interest \mathbf{U} , we employ a Top- k Target Attention mechanism:

$$\mathbf{U} = \text{Softmax}(\text{Top-}k(\mathbf{X}_{\text{tar}}\mathbf{X}_{\text{mix}}^T/\alpha, k)) \cdot \mathbf{X}_{\text{mix}}. \quad (6)$$

Unlike conventional target attention, the top- k target attention retains only the k items with the highest similarity. To ensure differentiability, we implement this by masking the non-top- k attention scores to $-\infty$ before the Softmax operation, allowing gradients to flow through the selected indices. This design draws inspiration from multi-task learning strategies that isolate parameters to prevent the seesaw phenomenon [10, 12, 14], ensuring that conflicting or weak interest signals do not dilute the final prediction.

3 Experiments

3.1 Experimental Setup

Datasets & Protocols. We evaluate FEDIN on three public datasets: **Tmall**¹, **Taobao**², and **Alipay**³. The datasets are partitioned into training, validation, and testing sets based on the global timeline. For evaluation, we employ AUC and **Group AUC (GAUC)** as the primary metrics [21].

¹<https://tianchi.aliyun.com/dataset/dataDetail?dataId=42>

²<https://tianchi.aliyun.com/dataset/dataDetail?dataId=649>

³<https://tianchi.aliyun.com/dataset/dataDetail?dataId=53>

Table 1: The overall performance of various methods on three publicly available recommendation datasets. Bold indicates the best performance, while underlined denotes the best performance of the baseline model. An asterisk (*) implies the improvements over the best baseline are statistically significant (p -value < 0.05).

Model	Tmall		Alipay		Taobao	
	GAUC	AUC	GAUC	AUC	GAUC	AUC
Sum Pooling	0.8978	0.8873	0.8557	0.8535	0.9345	0.9337
DIN	<u>0.9547</u>	<u>0.9518</u>	0.8897	0.8832	<u>0.9689</u>	<u>0.9664</u>
DIEN	0.9237	0.9157	0.8980	0.8953	0.9443	0.9442
SASRec	0.9183	0.9246	0.9238	<u>0.9312</u>	0.9583	0.9584
BERT4Rec	0.9103	0.9157	0.9179	0.9189	0.9523	0.9535
GRU4Rec	0.9210	0.9242	<u>0.9268</u>	0.9289	0.9618	0.9638
BST	0.9233	0.9269	0.9264	0.9285	0.9562	0.9576
DIFF	0.8513	0.8618	0.8962	0.8995	0.9463	0.9475
FEDIN (Ours)	0.9658*	0.9666*	0.9335*	0.9320*	0.9740*	0.9729*

Baselines. To comprehensively evaluate the effectiveness of our proposed FEDIN model, we conducted a thorough comparison with six state-of-the-art CTR prediction baseline models, including DIN [21], DIEN [22], SASRec [6], BERT4Rec [13], GRU4Rec [5], BST [1] and DIFF [7]. In addition, we utilize a Sum Pooling model as the base model, which transforms the list of user behavior embedding vectors into a fixed-length feature vector through sum pooling. Notably, while DIFF was originally designed for the next-item prediction task, we adapt it for the target-aware CTR task by replacing its ranking-based output layer with an MLP prediction head that incorporates the candidate item embedding.

Implementation Details. To ensure fair comparison, all models are implemented based on FuxiCTR [27]. We use the Adam optimizer with a learning rate of $5e^{-4}$ and a batch size of 2048. The embedding dimension is set to 32, and the maximum sequence length is fixed at 100.

3.2 Overall Performance

The quantitative results are presented in Table 1. FEDIN consistently outperforms the best-performing baselines across all three datasets, validating the efficacy of our frequency-enhanced approach. We observe that baseline models exhibit varying degrees of effectiveness depending on the dataset characteristics. For instance, attention-based models like DIN perform exceptionally well on Tmall and Taobao but struggle on Alipay, likely due to differences in data sparsity and behavior complexity. In contrast, FEDIN demonstrates superior robustness, achieving stable state-of-the-art performance in all scenarios. This stability can be attributed to the dual-branch architecture, where the frequency domain branch effectively compensates for the instability of time-domain features in sparse or noisy environments.

3.3 Ablation Study

To analyze the necessity of each component in the proposed model, we conducted ablation experiments. Table 2 presents the performance of FEDIN and its variants across three datasets. Overall, the removal of any component leads to a degradation in performance.

Table 2: Ablation analysis on three datasets. The best performance is denoted in bold.

Model	Tmall		Alipay		Taobao	
	GAUC	AUC	GAUC	AUC	GAUC	AUC
FEDIN	0.9658	0.9666	0.9335	0.9320	0.9740	0.9729
w/o Time-Domain Branch	0.9482	0.9498	0.9253	0.9246	0.9674	0.9669
w/o Freq-Domain Branch	0.9605	0.9618	0.9270	0.9263	0.9727	0.9715
w/o Freq-Domain TA	0.9614	0.9626	0.9327	0.9293	0.9732	0.9724
w/o Freq-Domain Scaling	0.9625	0.9636	0.9303	0.9296	0.9662	0.9655

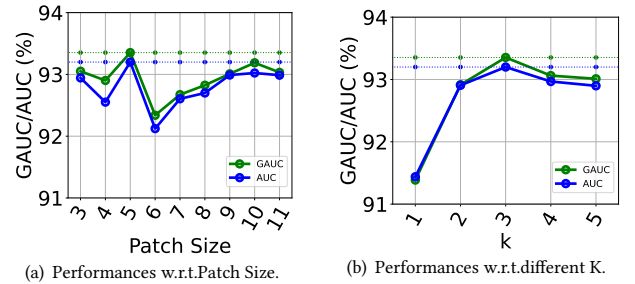


Figure 3: Hyperparameter analysis on the Taobao dataset.

Impact of Dual-Branch Architecture. We first evaluate the necessity of integrating both time and frequency domains. Removing the Time-Domain Branch (w/o Time-Domain Branch) leads to a significant performance drop, confirming that local sequential evolution is crucial for immediate interest prediction. Similarly, removing the Frequency-Domain Branch (w/o Freq-Domain Branch) also results in degradation. This proves that frequency-domain features capture complementary information—specifically global periodic patterns—that time-domain models alone fail to overlook.

Impact of Target-Aware Spectral Filtering. A key innovation of FEDIN is the incorporation of target items into the frequency analysis. In the w/o Freq-Domain TA variant, we replaced the target-aware spectrum generation with a standard learnable frequency filter. Specifically, we replaced the dynamic calculation of the attention score spectrum in Eqn. (3) with a static, learnable weight matrix. This matrix directly filters the user behavior spectrum, decoupling the frequency analysis from the target item. The decline in metrics validates our core hypothesis: simply analyzing the user’s behavior spectrum is insufficient. The target item acts as a necessary condition to activate specific resonance patterns, allowing the model to distinguish relevant signals from background noise.

Impact of Adaptive Scaling. The w/o Freq-Domain Scaling variant removes the adaptive gating mechanism. In this configuration, we omitted the scaling factor described in Eqn. (5) and directly used the unscaled output of the spectral filter as the branch output. The observed performance gap suggests that not all user-item interactions exhibit strong periodicity. The adaptive scaling allows the model to dynamically adjust the reliance on frequency features, prioritizing them only when clear spectral structures are detected, thereby enhancing model flexibility.

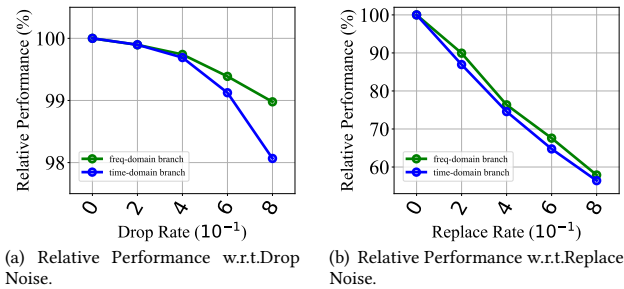


Figure 4: Noise resistance analysis on the Taobao dataset.

3.4 Hyperparameter Analysis

There are two hyperparameters of FEDIN that significantly impact performance: 1) the size of each patch in the time-domain branch, denoted as Patch Size; and 2) the number of candidate interest vectors to retain in the User Interest Aggregator, denoted as K . Fig. 3 illustrates the results of our experiments on public datasets to determine the optimal settings. For Patch Size, extreme values either fragment local context or cause information loss. For K , a balance is required between interest diversity and noise reduction; a small K limits coverage, while a large K introduces irrelevant signals and exacerbates the seesaw phenomenon.

3.5 Noise Resistance Ability Analysis

To further validate the robustness of our frequency-enhanced design, we conducted a synthetic noise resistance experiment. Since real-world interaction logs naturally suffer from sparsity and accidental clicks, we simulated two corresponding noise scenarios at varying corruption ratios ($\rho \in \{0.0, 0.2, 0.4, 0.6, 0.8\}$): (1) **Drop Noise**, which randomly deletes historical behaviors to mimic missing data, and (2) **Replace Noise**, which substitutes true behaviors with uniformly sampled random items to mimic exploratory or accidental clicks. We then compared the relative performance decay of the standalone Time-Domain Branch versus the Frequency-Domain Branch.

As shown in Fig. 4, the Frequency-Domain Branch demonstrates significantly higher stability as the noise rate ρ increases. In the time domain, deleting or replacing even a few items severely disrupts the local sequential context, often causing error propagation in standard RNNs or Transformers. Conversely, in the frequency domain, random noise typically manifests as dispersed, high-entropy fluctuations, whereas core user interests concentrate energy into distinct resonant peaks. By applying spectral filtering, FEDIN effectively attenuates these high-entropy noise components and reconstructs the underlying interest pattern from the corrupted raw sequence. This analysis confirms that integrating frequency-domain modeling significantly enhances the robustness of recommendation systems against real-world data imperfections.

4 Conclusions

This study explored methods for further applying frequency domain analysis techniques in CTR prediction. To facilitate this research, we propose FEDIN, which extracts frequency domain features by

incorporating a target attention mechanism and integrates both time domain and frequency domain analysis methods. This allows FEDIN to maintain stable performance across datasets with varying distributions. Comprehensive experiments conducted on three public datasets demonstrate FEDIN’s effectiveness.

Acknowledgments

We want to thank the anonymous reviewers and the meta-reviewer for their valuable comments and suggestions. This work is supported in part by the National Natural Science Foundation of China under Grants 62571298 and 624B2088.

References

- [1] Qiwei Chen, Huan Zhao, Wei Li, Pipei Huang, and Wenwu Ou. 2019. Behavior Sequence Transformer for E-commerce Recommendation in Alibaba. *CoRR* abs/1905.06874 (2019).
- [2] Tao Dai, Beiliang Wu, Peiyuan Liu, Naiqi Li, Jigang Bao, Yong Jiang, and Shu-Tao Xia. 2024. Periodicity Decoupling Framework for Long-term Series Forecasting. In *The Twelfth International Conference on Learning Representations, ICLR 2024*.
- [3] Xinyu Du, Huanhuan Yuan, Pengpeng Zhao, Jianfeng Qu, Fuzhen Zhuang, Guan-feng Liu, Yanchi Liu, and Victor S. Sheng. 2023. Frequency Enhanced Hybrid Attention Network for Sequential Recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 78–88.
- [4] Yufei Feng, Fuyu Lv, Weichen Shen, Menghan Wang, Fei Sun, Yu Zhu, and Keping Yang. 2019. Deep Session Interest Network for Click-Through Rate Prediction. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019*. 2301–2307.
- [5] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based Recommendations with Recurrent Neural Networks. In *4th International Conference on Learning Representations, ICLR 2016*.
- [6] Wang-Cheng Kang and Julian J. McAuley. 2018. Self-Attentive Sequential Recommendation. In *IEEE International Conference on Data Mining, ICDM 2018*. 197–206.
- [7] Hye-young Kim, Minjin Choi, Sunkyoung Lee, Ilwoong Baek, and Jongwuk Lee. 2025. DIFF: Dual Side-Information Filtering and Fusion for Sequential Recommendation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1624–1633.
- [8] Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. 2022. Reversible Instance Normalization for Accurate Time-Series Forecasting against Distribution Shift. In *The Tenth International Conference on Learning Representations, ICLR 2022*.
- [9] Chengkai Liu, Jianghao Lin, Jianling Wang, Hanzhou Liu, and James Caverlee. 2024. Mamba4Rec: Towards Efficient Sequential Recommendation with Selective State Space Models. *CoRR* abs/2403.03900 (2024).
- [10] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H. Chi. 2018. Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1930–1939.
- [11] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. 2023. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In *The Eleventh International Conference on Learning Representations, ICLR 2023*.
- [12] Liangcai Su, Junwei Pan, Ximei Wang, Xi Xiao, Shijie Quan, Xihua Chen, and Jie Jiang. 2024. STEM: Unleashing the Power of Embeddings for Multi-Task Recommendation. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024*. 9002–9010.
- [13] 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 2019*. 1441–1450.
- [14] Hongyan Tang, Junning Liu, Ming Zhao, and Xudong Gong. 2020. Progressive Layered Extraction (PLE): A Novel Multi-Task Learning (MTL) Model for Personalized Recommendations. In *Fourteenth ACM Conference on Recommender Systems, RecSys 2020*. 269–278.
- [15] Chiheb Trabelsi, Olexa Bilaniuk, Ying Zhang, Dmitriy Serdyuk, Sandeep Subramanian, Joao Felipe Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, and Christopher J. Pal. 2017. Deep complex networks. *arXiv preprint arXiv:1705.09792* (2017).
- [16] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems 30*. 5998–6008.
- [17] Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. 2023. Frequency-domain MLPs

- are More Effective Learners in Time Series Forecasting. In *Advances in Neural Information Processing Systems 36, NeurIPS 2023*.
- [18] Zhenrui Yue, Yueqi Wang, Zhankui He, Huimin Zeng, Julian J. McAuley, and Dong Wang. 2024. Linear Recurrent Units for Sequential Recommendation. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining, WSDM 2024*. 930–938.
- [19] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. 2023. Are Transformers Effective for Time Series Forecasting?. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023*. 11121–11128.
- [20] Qianru Zhang, Honggang Wen, Wei Yuan, Crystal Chen, Menglin Yang, Siu-Ming Yiu, and Hongzhi Yin. 2025. HMamba: Hyperbolic Mamba for Sequential Recommendation. *arXiv preprint arXiv:2505.09205 (2025)*.
- [21] Guorui Zhou et al. 2018. Deep Interest Network for Click-Through Rate Prediction. In *Proc. KDD*. 1059–1068.
- [22] Guorui Zhou, Na Mou, Ying Fan, Qi Pi, Weijie Bian, Chang Zhou, Xiaoqiang Zhu, and Kun Gai. 2019. Deep Interest Evolution Network for Click-Through Rate Prediction. In *The Thirty-Third AAAI Conference on Artificial Intelligence*. 5941–5948.
- [23] Haolin Zhou, Junwei Pan, Xinyi Zhou, Xihua Chen, Jie Jiang, Xiaofeng Gao, and Guihai Chen. 2024. Temporal Interest Network for User Response Prediction. In *Companion Proceedings of the ACM on Web Conference 2024, WWW 2024*. 413–422.
- [24] Kun Zhou, Hui Yu, Wayne Xin Zhao, and Ji-Rong Wen. 2022. Filter-enhanced MLP is All You Need for Sequential Recommendation. In *The ACM Web Conference 2022*. 2388–2399.
- [25] Tian Zhou, Ziqing Ma, Xue Wang, Qingsong Wen, Liang Sun, Tao Yao, Wotao Yin, and Rong Jin. 2022. FiLM: Frequency improved Legendre Memory Model for Long-term Time Series Forecasting. In *Advances in Neural Information Processing Systems 35, NeurIPS 2022*.
- [26] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. 2022. FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting. In *International Conference on Machine Learning, ICML 2022 (Proceedings of Machine Learning Research, Vol. 162)*. 27268–27286.
- [27] Jieming Zhu, Jinyang Liu, Shuai Yang, Qi Zhang, and Xiuqiang He. 2021. Open Benchmarking for Click-Through Rate Prediction. In *The 30th ACM International Conference on Information and Knowledge Management, CIKM '21*. 2759–2769.