Efficient Noise-reducing Neural Network for Cross-Domain Sequential Recommendation

Kaiwei Xu[1], Yongquan Fan[1(✉)], and Jing Tang[2]

[1] School of Computer and Software Engineering, Xihua University, Chengdu, China
xkw@stu.xhu.edu.cn, fyqxhu@gmail.com
[2] Southwest Petroleum University, Chengdu, China
tangjing@swpu.edu.cn

Abstract. Cross-domain Sequential Recommendation (CSR) is a developing task that examines how users' sequential patterns evolve by analyzing their interactions in multiple domains. Existing CSR methods still face the following challenges: First, users' interaction behaviors across different domains are highly complex, diverse, and challenging to predict. Second, these methods often overlook the impact of noise during the knowledge transfer process between domains. To address these challenges, we propose a novel Efficient Noise-reducing Neural Network for Cross-Domain Sequential Recommendation (ENNR). Specifically, we first employ filter encoders from a frequency domain perspective to encode the sequence data in single-domain and cross-domain, which captures the complex sequential patterns of users and improves the computational efficiency of the model. In addition, to reduce the impact of noise on model performance, we design sparse self-attention mechanisms that aim to avoid the adverse effects of the noise item at the end of the sequence, while ensuring minimal weight is assigned to randomly occurring noise items in the sequences. Through the above methods, we achieve accurate knowledge transfer, improving the overall performance of our model in CSR tasks. The final experiments on four public datasets show that the proposed ENNR surpasses state-of-the-art baselines in accuracy and efficiency.

Keywords: Recommendation systems · Cross-domain sequential recommendation · Filtering algorithm · Sparse attention mechanism

1 Introduction

In the era of rapidly advancing information technology, recommender systems [18,16] have become a crucial solution to the issue of information overload. Particularly in environments aimed at meeting user needs, the role of recommender systems is increasingly prominent. With the diversification of recommendation environments and the expansion of the recommendation scale, the issues of data sparsity and cold start have intensified. To address this challenge, researchers have proposed Cross-Domain Sequential Recommendation (CSR) algorithms [17,12,7], which enhance recommendation performance by integrating data from different domains. The pioneering work in CSR is PiNet [14], which transfers sequential representations from a single domain to others through a gated transfer module. Recently, with advancements in neural networks, Graph Neural Network (GNN)-based methods [1,6] and attention-based methods [10,13] have become focal points of research. For instance, *Cao et al.* [1] utilizes GNN to capture high-order collaborations between items and design a contrastive loss function for dynamic representations in and between sequences, enabling precise modeling of user preferences. *Lin et al.* [11] introduces a hybrid self-attention framework to better facilitate information transfer across different domains.

Although some progress has been made, there are still some challenges in CSR. This is primarily due to two factors. Firstly, existing methods primarily focus on modeling sequential patterns across various domains in the temporal dimension. However, users' behaviors are influenced by the temporal dimension, exhibiting complex and variable characteristics, which makes prediction more difficult. Take Amazon's online shopping platform as an example, in the beauty domain, affected by emotions or popular fashions, users' behaviors show high randomness and short-term variability, making it difficult to determine clear purchase intentions. In contrast, in the electronics domain, users may thoroughly investigate item specifications and reviews, but their purchasing decisions are affected by time-sensitive factors such as technology updates and price fluctuations. These behavioral differences across domains make it challenging to capture users' true preferences and needs by relying solely on the temporal dimension, and reduce the efficiency of model training [20]. Second, user interaction sequences often contain noise, which negatively impacts the accuracy of prediction. Specifically, users may occasionally click on items unrelated to their interests; these noise items appear randomly in the sequence and may even occur as the last item, thereby adversely affecting the prediction of the next item [18]. Thus, how to effectively deal with noise has become a critical issue in CSR.

To address the aforementioned challenges, we propose a novel model called Efficient Noise-reducing Neural Network for Cross-Domain Recommendation (ENNR). For the first challenge, we shift our perspective to the frequency domain and design two filter encoders to encode user sequences in single-domain and cross-domain. These encoders can effectively capture complex user sequential patterns while improving computational efficiency by replacing complex convolution operations in the time domain with fast Fourier transform (FFT) [15]. To tackle the second challenge, we develop two sparse self-attention mechanisms for single-domain and cross-domain sequences. Specifically, we avoid the adverse impact of noise from the last item in the sequence on model prediction by learning a target representation designed to reduce noise inclusion. At the same time, we employ an adaptive sparse transformation function within the sparse self-attention mechanisms to minimize the weight assigned to randomly occurring noise items, which reduces how noise in the sequence interferes with the model. We conduct experiments on four public datasets to evaluate the performance of the model. Extensive results show that our model ENNR achieves significant improvements over the state-of-the-art baselines while maintaining efficiency.

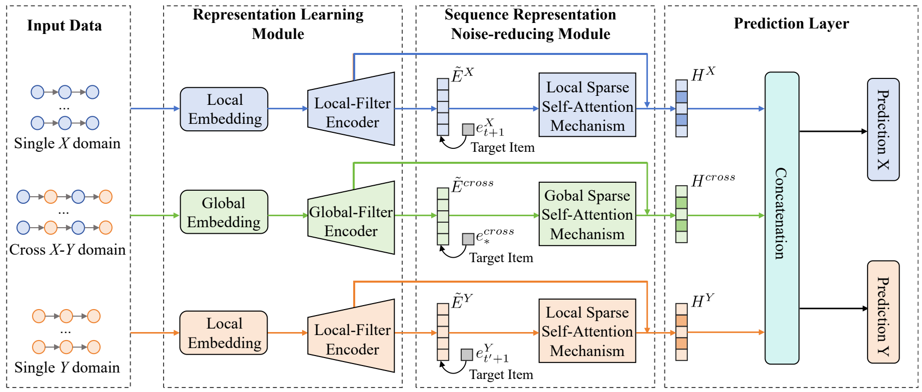

Fig. 1. The overview of ENNR.

2 Problem Formulation

Considering two distinct domains X and Y, we have a set of overlapping users across these domains, denoted as $\mathcal{U} = \{u_1, u_2, \ldots, u_k\}$, where k represents the number of overlapping users. For each user $u \in \mathcal{U}$, let $\mathcal{X}$ and $\mathcal{Y}$ denote the sets of items in the domains X and Y, respectively. Specifically, suppose $x_t \in \mathcal{X}$ and $y_{t'} \in \mathcal{Y}$ represent the t-th and t'-th items that the user has interacted with in their respective single domains. The sequences of their past interactions can then be expressed as $S^X = (x_1, x_2, \ldots, x_t)$ and $S^Y = (y_1, y_2, \ldots, y_{t'})$. By chronologically merging these sequences, we obtain the cross-domain interaction sequence $S^{\text{cross}} = (y_1, x_1, x_2, \ldots, y_{t'}, \ldots, x_t)$. Based on the sequences S^X, S^Y, and S^{cross}, CSR seeks to predict the forthcoming item a user is likely to consume within domains X and Y.

3 Method

The main structure of our ENNR is shown in Figure 1, which describes the modules in detail in this section.

3.1 Representation Learning Module

In this section, embedding matrices for items in the single-domain and cross-domain are first generated. These embeddings are then processed using local and global filter encoders to model users' sequential patterns in each domain.

Embedding Initialization Layer. We define three parameter matrices: $I^X \in R^{|\mathcal{X}| \times d}$, $I^Y \in R^{|\mathcal{Y}| \times d}$ and $I \in R^{|\mathcal{X}+\mathcal{Y}| \times d}$ to initialize the item representations for the single-domain and cross-domain, where d denotes the dimension of the embedding.

Here, we use $(x_1, x_2, \ldots, x_t)$ and $(y_1, y_2, \ldots, y_{t'})$ to represent the interaction sequences in the corresponding single-domain, and $(x_1, y_1, y_2, \ldots, x_t, \ldots, y_{t'})$ for the cross-domain. Moreover, to capture the sequential order information, we integrate learnable positional embeddings into the item embeddings:

$$\begin{aligned} E^X &= \left[I^X_{x_1} + P^X_{x_1}, I^X_{x_2} + P^X_{x_2}, \ldots, I^X_{x_t} + P^X_{x_t}\right] \\ E^Y &= \left[I^Y_{y_1} + P^Y_{y_1}, I^Y_{y_2} + P^Y_{y_2}, \ldots, I^Y_{y_{t'}} + P^Y_{y_{t'}}\right] \end{aligned} \tag{1}$$

$$E^{cross} = \left[I_{x_1} + P_{x_1}, I_{y_1} + P_{y_1}, \ldots, I_{x_t} + P_{x_t}, \ldots, I_{y_{t'}} + P_{y_{t'}}\right] \tag{2}$$

where E^X and $E^Y \in R^{L\times d}$ represent the local embeddings for single domains X and Y, respectively, and $E^{cross} \in R^{L\times d}$ denote the global embedding for the cross-domain $X - Y$. Here, L is the maximum length of the interaction sequence, P^X, P^Y and $P \in R^{L\times d}$ are the learnable positional embeddings.

Local Filter Encoder in Single Domain. Filtering algorithms have recently emerged as a promising modeling technique in recommendation systems [20,3]. Leveraging the Fast Fourier Transform (FFT) with its $O(N \log N)$ complexity allows for the effective extraction of users' complex sequential patterns and improves model training efficiency. Therefore, we design a local filtering encoder to capture single-domain sequential patterns. Specifically, we first employ a one-dimensional FFT $\mathcal{F}(\cdot)$ to convert the input data into frequency domain representations Z^*. Next, a learnable filter W^* is applied to modulate the spectrum of the frequency domain representation Z^*. Finally, the modulated spectrum $\tilde{Z}^*$ is transformed back into the time domain representation $\tilde{E}^*$ via a one-dimensional inverse FFT (IFFT). This process can be expressed as follows:

$$\begin{aligned} &Z^* = \mathcal{F}(E^*) \in \mathbb{C}^{L\times d} \\ &\tilde{Z}^* = W^* \odot Z^* \\ &\tilde{E}^* \leftarrow \mathcal{F}^{-1}\left(\tilde{Z}^*\right) \in \mathbb{R}^{L\times d} \end{aligned} \tag{3}$$

where $\odot$ is the element-wise multiplication, and symbol $*$ can be X or Y.

Global Filter Encoder in Cross Domain. Compared with single-domain data, cross-domain data is full of different types of interactions. It may be difficult to extract relevant information in cross-domain data by relying solely on the local filter encoder provided by Eq (3). Therefore, it is crucial to develop an appropriate method to capture users' cross-domain sequential patterns. Inspired by [5], we propose a global filter encoder based on a two-layer multi-layer perceptron (MLP) structure to achieve this as follows:

$$\begin{aligned} &Z^{cross} = \mathcal{F}(E^{cross}) \in \mathbb{C}^{L\times d} \\ &\tilde{Z}^{cross} = SoftShrink\left(W_2\left(ReLU\left(W_1 Z^{cross} + b_1\right)\right) + b_2\right) \\ &\tilde{E}^{cross} \leftarrow \mathcal{F}^{-1}\left(\tilde{Z}^{cross}\right) \in \mathbb{R}^{L\times d} \end{aligned} \tag{4}$$

where W_1, $W_2 \in \mathbb{C}^{d\times d}$ denote the weight matrices used to process different feature, and $SoftShrink(\cdot)$ is an activation function defined as $SoftShrink(x) =$

$sign(x) \cdot \max(|x| - \lambda, 0)$, which retains input values exceeding the threshold parameter λ, while setting values less than or equal to the threshold to zero, thereby filtering out unimportant features. In this way, our global filter encoder can effectively extract relevant information in cross-domain data and thus accurately capture users' sequential patterns.

3.2 Sequence Representation Noise-reducing Module

After obtaining single-domain and cross-domain sequence representations $\tilde{E}^X$, $\tilde{E}^Y$ and $\tilde{E}^{cross}$, we specifically designed local and global sparse attention mechanisms for them to reduce the impact of noise in the sequence on model performance.

Local Sparse Self-Attention in Single Domain. In attention mechanisms, the traditional softmax function allocates probabilities to every item in the sequence without considering their significance to the user's intent [2]. To end this, we propose a local sparse attention mechanism to learn noise-reduced representations in single-domain sequences $\tilde{E}^X$ ($\tilde{E}^Y$) to enhance the accuracy predictions. First, we add a blank item without any sequence information at the end of the sequence to learn a target representation that reduces noise inclusion, mitigating the noise interference of the last item of the sequence.

$$\hat{E}^X = Concat(\tilde{E}^X, e^X_{t+1}), \quad \hat{E}^Y = Concat(\tilde{E}^Y, e^Y_{t'+1}) \tag{5}$$

where $e^X_{t+1}, e^Y_{t'+1} \in \mathbb{R}^d$ represent the additional target items, and $\hat{E}^X, \hat{E}^Y \in \mathbb{R}^{(L+1)\times d}$ are the updated sequence representations. Subsequently, taking $\hat{E}^X$ as an example, we set $K_s = V_s = \hat{E}^X$. However, to enable the self-attention mechanism to accurately distinguish between Q_s, K_s, and V_s, we perform a linear projection on Q_s, which enhances the model's flexibility in processing information. The process is defined as follows:

$$Q_s = ReLU(\hat{E}^X W_Q) \tag{6}$$

where $W_Q \in \mathbb{R}^{d\times d}$ denote the weight matrix, then we compute the parameter α_s used to adjust the sparsity of the entmax function for single-domain,

$$\alpha_s = Sigmoid(W_{\alpha_s} e^X_{t+1} + b_{\alpha_s}) + 1 \tag{7}$$

where $W_{\alpha_s} \in \mathbb{R}^{1\times d}$ represents the weight matrix, $b_{\alpha_s} \in \mathbb{R}^d$ represents the bias, and the **l**ocal **s**parse **a**ttention (LSA) mechanism is defined as follows,

$$LSA(\hat{E}^X) = SAtten\left(Q_s, K_s, V_s\right) \tag{8}$$

$$SAtten(Q_s, K_s, V_s) = \alpha_s{-}entmax\left(\frac{Q_s K_s^T}{\sqrt{d}}\right) V_s \tag{9}$$

where Q_s, K_s, and V_s are the query matrix, key matrix, and value matrix, respectively. Afterward, we also utilize residual connections, dropout, and layer normalization to obtain the final output,

$$\tilde{E}^X_s, \tilde{e}^X_{t+1} = LSA(\hat{E}^X) \tag{10}$$

where $\tilde{E}_s^X \in \mathbb{R}^{L\times d}$ denotes the new sequence representation for single-domain X, and $\tilde{e}_{t+1}^X \in \mathbb{R}^d$ is the corresponding learned target representation. Similarly, for single-domain Y, we obtain the new sequence representation $\tilde{E}_s^Y$ and the learned target representation $\tilde{e}_{t'+1}^Y$.

Gobal Sparse Self-Attention in Cross Domain. Although the local sparse attention mechanism can reduce noise in the single-domain, in the cross-domain, different domains contain varying numbers of interaction items, and the changes of noise items in the sequence are more random and uncertain. Therefore, we propose a global sparse attention mechanism. Specifically, we first represent the blank item $\bar{e}_*^{cross}$ in the cross-domain as a Gaussian distribution [4] that can capture the uncertainty in the item representation and thus provide richer information,

$$\begin{aligned} \mu^{cross} &= LeakyReLU(\bar{e}_*^{cross} W_\mu^{cross}) \\ \sigma^{cross} &= Softplus(\bar{e}_*^{cross} W_\sigma^{cross}) \\ e_*^{cross} &= \mu^{cross} + \sigma^{cross} \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \end{aligned} \tag{11}$$

where W_μ^{cross} and $W_\sigma^{cross} \in \mathbb{R}^{d\times d}$ are weight matrices and I is identity matrix. μ^{cross} is used to capture the central tendency of user-item interactions, while σ^{cross} reflects the variability of these interactions, enabling the model to capture the relationship between users and items more effectively. Note that considering the cross-domain, the symbol * can be $t+1$ or $t'+1$. In this way, we obtain the target item representing e_*^{cross}, and then we integrate it into the cross-domain sequence representation:

$$\hat{E}^{cross} = Concat(\tilde{E}^{cross}, e_*^{cross}) \tag{12}$$

where $\hat{E}^{cross} \in \mathbb{R}^{(L+1)\times d}$. Next, we compute the entmax function's sparsity parameter α_c for cross-domain as:

$$\alpha_c = Sigmoid(W_{\alpha_c} e_*^{cross} + b_{\alpha_c}) + 1 \tag{13}$$

where $W_{\alpha_c} \in \mathbb{R}^{1\times d}$ represents the weight matrix, $b_{\alpha_c} \in \mathbb{R}^d$ represents the bias. subsequently, we use the **g**obal **s**parse **a**ttention (GSA) mechanism to reduce the influence of noise items and improve the accuracy and robustness of the cross-domain sequence representation, which is defined as follows:

$$GSA(\hat{E}^{cross}) = SAtten\left(Q_c, K_c, V_c\right) \tag{14}$$

$$SAtten(Q_c, K_c, V_c) = \alpha_c\text{--}entmax\left(\frac{Q_c K_c^T}{\sqrt{d}}\right) V_c \tag{15}$$

where Q_c, K_c, V_c represent the query, key, and value matrices. Note that, following a similar approach to Eq.(6), we set $K_c = V_c = \hat{E}^{cross}$, and perform a linear operation on Q_c. Finally, we obtain the final output by residual connections, dropout, and layer normalization,

$$\tilde{E}_c^{cross}, \tilde{e}_*^{cross} = GSA(\hat{E}^{cross}) \tag{16}$$

Where $\tilde{E}_c^{cross} \in \mathbb{R}^{L\times d}$ denotes the new sequence representation for cross-domain, and $\tilde{e}_*^{cross} \in \mathbb{R}^d$ represents the learned target representation.

3.3 Prediction and Optimization

We generate the final sequence representation by combining information from two perspectives: filter encoder and sparse self-attention. Take single-domain X as an example,

$$H^X = Selu(W_3(\tilde{E}^X + \tilde{E}_s^X + \tilde{e}_{t+1}^X) + b_3) \tag{17}$$

where $W_3 \in \mathbb{R}^{d \times d}$ denote the weight matrix and $b_3 \in \mathbb{R}^d$ is the bias. After obtaining the final sequence representations H^X, H^Y, and H^{cross} for both the single-domain and cross-domain, we follow previous works [19], which leverage information from both domains to compute the recommendation probabilities,

$$\begin{aligned} P(x_{t+1} \mid S^X, S^{cross}) &= softmax\left(W_{X,cross} \cdot \left[H^X,\ H^{cross}\right]^T + b_{X,cross}\right) \\ P(y_{t'+1} \mid S^Y, S^{cross}) &= softmax\left(W_{Y,cross} \cdot \left[H^Y,\ H^{cross}\right]^T + b_{Y,cross}\right) \end{aligned} \tag{18}$$

where $W_{X,cross}, W_{Y,cross}$ are the weight matrices for all items, and $b_{X,cross}, b_{Y,cross}$ are the biases. Subsequently, we apply the negative log-likelihood loss function to optimize the model as follows:

$$\begin{aligned} \mathcal{L}_X &= -\frac{1}{|\mathcal{S}|} \sum_{S^X \in \mathcal{S}} \sum_{x_t \in S^X} \log P(x_{t+1} \mid S^X, S^{cross}), \\ \mathcal{L}_Y &= -\frac{1}{|\mathcal{S}|} \sum_{S^Y \in \mathcal{S}} \sum_{y_{t'} \in S^Y} \log P(y_{t'+1} \mid S^Y, S^{cross}). \end{aligned} \tag{19}$$

where $\mathcal{S}$ represents the training sequences in both domains.

4 Experiments

4.1 Experimental Setting

Datasets. In this section, we evaluate the performance of ENNR by conducting experiments on four different domains of Amazon public datasets[3]: "Cloth & Sport" and "Phone & Elec". In each dataset, we select overlapping users present in both domains and remove users and their associated items that have fewer than 5 interactions within these domains. Detailed information about these datasets is outlined below (Table 1).

Compared Methods and Settings. We compare the proposed ENNR with the following baseline models: 1) Single-domain sequential recommendations: GRU4REC [8], SASRec [9], and FMLP-Rec [20]. 2) Cross-domain sequential recommendations: PiNet [14], DA-GCN [6], DASL [10], C^{2}DSR [1], and MAN [11]. To train the model, we set the batch size to 256 and the embedding dimension to 64 for experimentation. For the global filtering encoder of ENNR, the parameter λ is set to 0.01. The evaluation metrics include HR@k and NDCG@k, with k values of 5 and 10, applied to each dataset.

[3] http://jmcauley.ucsd.edu/data/amazon/

Table 1. Experimental datasets.

Dataset	**AMAZON**			
Domain	**Cloth**	**Sport**	**Phone**	**Elec**
Users	27,519	107,984	41,829	27,328
Items	9,481	40,460	17,943	12,655
Interactions	161,010	851,553	194,121	170,426
#Overlap Users	16,337		7,857	
Avg.length	4.39	7.58	4.53	6.19

4.2 Performance Comparisons

Overall Comparisons. In Table 2, we evaluate the proposed model's performance in four datasets using HR@[5, 10] and NDCG@[5, 10] metrics, comparing them with baseline models. CSR methods (e.g., DA-GCN, DASL) generally outperform single-domain methods (e.g., GRU4REC, SASRec), highlighting the importance of cross-domain data in improving recommendations. Additionally, GNN-based models (e.g., DA-GCN, C^{2}DSR) leverage graph structures to capture high-order relationships, enhancing feature extraction and accuracy. Although the attention-based MAN performs best among baselines, its performance is still affected by domain noise, making it inferior to our model. Our ENNR model, with its filter encoders and sparse self-attention mechanisms, effectively reduces noise, achieving superior performance in all tasks.

Table 2. Experimental results of recommendation methods on evaluation datasets are shown.

Datasets	**Cloth & Sport**								**Phone & Elec**							
Domain	**Cloth**				**Sport**				**Phone**				**Elec**			
Metric	**H@5**	**H@10**	**N@5**	**N@10**	**H@5**	**H@10**	**N@5**	**N@10**	**H@5**	**H@10**	**N@5**	**N@10**	**H@5**	**H@10**	**N@5**	**N@10**
GRU4REC	5.93	6.67	2.84	3.51	8.59	9.87	4.14	5.12	11.45	12.62	5.51	7.14	18.65	19.34	8.78	10.45
SASRec	6.26	6.94	2.90	3.59	8.74	10.09	4.23	5.15	11.66	12.98	5.74	7.22	19.16	20.13	9.05	11.69
FMLP-Rec	6.44	7.13	3.08	3.72	9.09	10.51	4.31	5.30	12.59	13.92	5.98	7.59	19.33	20.24	9.17	11.83
PINet	5.23	6.44	2.45	3.11	7.88	9.24	3.88	4.95	11.68	12.87	5.77	7.35	18.87	19.59	8.93	10.63
DA-GCN	6.31	7.05	2.76	3.54	8.81	10.16	4.29	5.26	11.92	13.29	5.85	7.38	19.29	20.21	9.10	11.77
DASL	6.29	7.02	2.95	3.62	8.83	10.13	4.17	5.19	11.90	13.25	5.83	7.41	19.25	20.17	9.15	11.86
C^{2}DSR	6.52	7.19	3.14	3.77	9.13	10.56	4.36	5.34	13.38	14.66	6.41	8.17	19.60	20.56	9.08	11.75
MAN	<u>6.84</u>	<u>7.51</u>	<u>3.36</u>	<u>4.04</u>	<u>10.09</u>	<u>11.75</u>	<u>4.75</u>	<u>5.82</u>	<u>14.35</u>	<u>15.58</u>	<u>6.68</u>	<u>8.36</u>	<u>19.72</u>	<u>20.87</u>	<u>9.26</u>	<u>12.05</u>
ENNR	**7.38**	**8.01**	**3.62**	**4.39**	**11.22**	**12.95**	**5.41**	**6.53**	**15.06**	**16.23**	**7.06**	**8.82**	**19.95**	**21.14**	**9.67**	**12.34**
Improv.	7.89%	6.65%	7.73%	8.66%	11.19%	10.21%	13.89%	12.19%	4.95%	4.17%	5.69%	5.50%	1.16%	1.29%	4.42%	2.41%

Model Training Efficiency. To evaluate the training time consumption of the models, we vary the proportion of input data on the AMAZON dataset (ranging from 0.2 to 1.0) and conduct a series of comparative experiments against two highly competitive baseline models, C^{2}DSR and MAN. As shown in Figure. 2(a) and (b), ENNR exhibits shorter training times compared to C^{2}DSR and MAN. These findings further demonstrate that our proposed ENNR not only offers higher training efficiency and superior performance but also scales effectively to large-scale datasets.

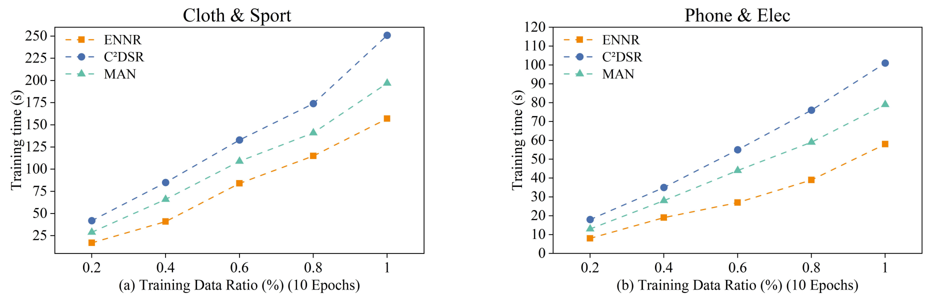

Fig. 2. Time consumption of ENNR compared with C^{2}DSR and MAN.

Table 3. Results of ablation study.

Datasets	**Cloth & Sport**							
Domain	**Cloth**				**Sport**			
Metric	**H@5**	**H@10**	**N@5**	**N@10**	**H@5**	**H@10**	**N@5**	**N@10**
w/o.mg	6.93	7.72	3.19	3.97	10.67	12.15	4.64	5.85
w/o.satten	5.96	6.51	2.94	3.44	9.81	11.09	3.77	4.90
w/o.lf	6.44	7.13	3.05	3.68	10.06	11.62	4.03	5.18
ENNR	**7.38**	**8.01**	**3.62**	**4.39**	**11.22**	**12.95**	**5.41**	**6.53**

4.3 Ablation Study

To analyze the impact of each ENNR component, we conduct ablation experiments with three variants: *w/o.mg* (global filter encoder without MLP and Gaussian representation), *w/o.satten* (removes sparse attention mechanisms), and *w/o.lf* (removes filter encoders). Results in Table 3 show that *w/o.mg* performs poorly, highlighting the importance of strategies to capture cross-domain interactions. *w/o.satten* performs worst, demonstrating the crucial role of sparse attention in reducing noise. *w/o.lf* shows that filter encoders from the frequency domain effectively capture complex patterns, improving prediction accuracy.

5 Conclusion

We propose an ENNR model to address noise in CSR. We design filter encoders to generate sequence embeddings from single-domain and cross-domain sequences, effectively capturing users' complex patterns and improving training efficiency. To reduce noise impact, we introduce sparse attention mechanisms that minimize the weight of noise items in sequences, preventing them from affecting prediction accuracy. Experimental results on four datasets show that our model outperforms the latest CSR methods.

Acknowledgments. This work was supported by the Sichuan Science and Technology Program (No. 2023YFQ0044, 2024ZHCG001) and the Postgraduate Course Teaching Team Project of Xihua University (No. YTD202303).

References

1. Cao, J., Cong, X., Sheng, J., Liu, T., Wang, B.: Contrastive cross-domain sequential recommendation. In: CIKM. pp. 138–147 (2022)
2. Correia, G.M., Niculae, V., Martins, A.F.: Adaptively sparse transformers. In: EMNLP (2019)
3. Du, X., Yuan, H., Zhao, P., Qu, J., Zhuang, F., Liu, G., Liu, Y., Sheng, V.S.: Frequency enhanced hybrid attention network for sequential recommendation. In: SIGIR. pp. 78–88 (2023)
4. Fan, Z., Liu, Z., Wang, Y., Wang, A., Nazari, Z., Zheng, L., Peng, H., Yu, P.S.: Sequential recommendation via stochastic self-attention. In: WWW. pp. 2036–2047 (2022)
5. Guibas, J., Mardani, M., Li, Z., Tao, A., Anandkumar, A., Catanzaro, B.: Adaptive fourier neural operators: Efficient token mixers for transformers. In: ICLR (2022)
6. Guo, L., Tang, L., Chen, T., Zhu, L., Nguyen, Q.V.H., Yin, H.: Da-gcn: a domain-aware attentive graph convolution network for shared-account cross-domain sequential recommendation. In: IJCAI. pp. 2483–2489 (2021)
7. Guo, L., Zhang, J., Chen, T., Wang, X., Yin, H.: Reinforcement learning-enhanced shared-account cross-domain sequential recommendation. TKDE **35**(7), 7397–7411 (2022)
8. Hidasi, B., Karatzoglou, A., Baltrunas, L., Tikk, D.: Session-based recommendations with recurrent neural networks. In: ICLR (2016)
9. Kang, W.C., McAuley, J.: Self-attentive sequential recommendation. In: ICDM. pp. 197–206 (2018)
10. Li, P., Jiang, Z., Que, M., Hu, Y., Tuzhilin, A.: Dual attentive sequential learning for cross-domain click-through rate prediction. In: KDD. pp. 3172–3180 (2021)
11. Lin, G., Gao, C., Zheng, Y., Chang, J., Niu, Y., Song, Y., Gai, K., Li, Z., Jin, D., Li, Y., et al.: Mixed attention network for cross-domain sequential recommendation. In: WSDM. pp. 405–413 (2024)
12. Liu, W., Zheng, X., Chen, C., Su, J., Liao, X., Hu, M., Tan, Y.: Joint internal multi-interest exploration and external domain alignment for cross domain sequential recommendation. In: WWW. pp. 383–394 (2023)
13. Ma, H., Xie, R., Meng, L., Chen, X., Zhang, X., Lin, L., Zhou, J.: Triple sequence learning for cross-domain recommendation. TOIS **42**(4), 1–29 (2024)
14. Ma, M., Ren, P., Lin, Y., Chen, Z., Ma, J., Rijke, M.d.: π-net: A parallel information-sharing network for shared-account cross-domain sequential recommendations. In: SIGIR. pp. 685–694 (2019)
15. Nussbaumer, H.J., Nussbaumer, H.J.: The fast Fourier transform. Springer (1982)
16. Qiao, S., Zhou, W., Wen, J., Zhang, H., Gao, M.: Bi-channel multiple sparse graph attention networks for session-based recommendation. In: CIKM. pp. 2075–2084 (2023)
17. Xu, W., Wu, Q., Wang, R., Ha, M., Ma, Q., Chen, L., Han, B., Yan, J.: Rethinking cross-domain sequential recommendation under open-world assumptions. In: WWW. pp. 3173–3184 (2024)
18. Yuan, J., Song, Z., Sun, M., Wang, X., Zhao, W.X.: Dual sparse attention network for session-based recommendation. In: AAAI. vol. 35, pp. 4635–4643 (2021)
19. Zhang, J., Duan, H., Guo, L., Xu, L., Wang, X.: Towards lightweight cross-domain sequential recommendation via external attention-enhanced graph convolution network. In: DASFAA. pp. 205–220. Springer (2023)
20. Zhou, K., Yu, H., Zhao, W.X., Wen, J.R.: Filter-enhanced mlp is all you need for sequential recommendation. In: WWW. pp. 2388–2399 (2022)