

Gating-adapted Wavelet Multiresolution Analysis for Exposure Sequence Modeling in CTR prediction

Xiaoxiao Xu, Zhiwei Fang, Qian Yu, Ruoran Huang,
Chaosheng Fan, Yong Li, Yang He, Changping Peng, Zhangang Lin, Jingping Shao
Business Growth BU, JD.com

Beijing

{xuxiaoxiao1,fangzhiwei2,yuqian81,huangruoran1}@jd.com
{fanchaosheng1,liyong5,landy,pengchangping,linzhangang,shaojingping}@jd.com

ABSTRACT

The exposure sequence is being actively studied for user interest modeling in Click-Through Rate (CTR) prediction. However, the existing methods for exposure sequence modeling bring extensive computational burden and neglect noise problems, resulting in an excessively latency and the limited performance in online recommenders. In this paper, we propose to address the high latency and noise problems via **Gating-adapted wavelet multiresolution analysis** (Gama), which can effectively denoise the extremely long exposure sequence and adaptively capture the implied multi-dimension user interest with linear computational complexity. This is the first attempt to integrate non-parametric multiresolution analysis technique into deep neural networks to model user exposure sequence. Extensive experiments on large scale benchmark dataset and real production dataset confirm the effectiveness of Gama for exposure sequence modeling, especially in cold-start scenarios. Benefited from its low latency and high effectiveness, Gama has been deployed in our real large-scale industrial recommender, successfully serving over hundreds of millions users.

CCS CONCEPTS

• **Information systems** → **Retrieval models and ranking**.

KEYWORDS

CTR prediction, Exposure sequence, Multiresolution analysis

ACM Reference Format:

Xiaoxiao Xu, Zhiwei Fang, Qian Yu, Ruoran Huang, Chaosheng Fan, Yong Li, Yang He, Changping Peng, Zhangang Lin, Jingping Shao. 2022. Gating-adapted Wavelet Multiresolution Analysis for Exposure Sequence Modeling in CTR prediction. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22)*, July 11–15, 2022, Madrid, Spain. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3477495.3531771>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGIR '22, July 11–15, 2022, Madrid, Spain

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-8732-3/22/07... \$15.00

<https://doi.org/10.1145/3477495.3531771>

1 INTRODUCTION

To present the most attractive items for different users, the Click-Through Rate (CTR) prediction algorithms in modern recommenders are always equipped with user interest modeling [2, 8, 12, 13]. User behavior sequence, such as click and buy, are commonly utilized as the information sources for extracting user interest. Recently, some methods also incorporate the sequential exposure data to improve user interest modeling [5, 11]. In e-commerce recommender, an exposure conveys the information about the corresponding product to the user, such as its appearance, price, selling points, etc. Thus, the user can be impacted by the abundant exposures. That is, we can extract the implicit interest from user exposure histories.

High latency and noise become the major constraints for utilizing exposure sequence in online recommenders. Different from the user behavior data, two major characteristics of the exposure data need to be concerned: 1) The sequential exposure data is awfully denser than user behavior, with a 1:20 ratio approximately in length. This density issue raises a stricter requirement for the efficiency of recommendation systems to exploit exposure data. 2) The exposure data is noisy, since not all the information can be received by the user in the dense exposure sequence. However, detail information about whether the user actually saw the product, and if so, how long the engagement lasted, are not available due to the device constraint and privacy issue. Consequently, detail information's absence in exposure data brings noise for exposure sequence modeling. Generally, the isolated exposures are regarded as noise which have no connection with any aspect concerned by the user, while too much noise will reduce the effect of user interest modeling.

In this paper, we propose **Gating-adapted wavelet multi-resolution analysis** (Gama) to address the high latency and noise problems when modeling exposure sequence. Specifically, Gama regards the exposure sequence as a sampled time-varying signal, and decomposes the signal into components with different frequencies. Thus the noise can be reduced by abandoning the high-frequency components. Besides, Gama utilizes the Interest Gate Net to adapt wavelet MRA to further denoise and boost the performance by extracting the most important multi-dimension user interest. Specifically, Interest Gate Net reweights the multiple components with different frequencies adaptive to user behavior histories. Gama enjoys linear computational complexity in the critical exposure signals learning stage which results in low inference latency. Incorporating Gama in CTR prediction model can help to enhance user interest modeling by introducing exposure sequence, especially for the cold-start users with sparse behaviors. We valid the effectiveness and low latency of

Gama through extensive experiments on public dataset and in real production environment, and successfully apply it in a real-world large scale display advertising system whose latency requirements are extremely rigorous.

2 THE PROPOSED METHOD

2.1 Preliminary

Given the user, the candidate item and the contexts in an exposure scenario, CTR prediction is to predict the conditional probability of a click event. Thus the user interest modeling module is crucial in CTR prediction. Traditionally, user interest is learned from the user behavior histories, with specific modules equipped based on attention mechanism or RNN.

In this paper, we focus on enhancing user interest modeling by introducing exposure histories. Similar to user behavior histories, user exposure histories can be fully recorded in the form of sequences over time. Let S^u denotes the ordered exposure histories of user u :

$$S^u = [x_1^u, x_2^u, \dots, x_N^u] \quad (1)$$

where x_i^u denotes the item ID of user u at the i -th exposure, and N is the length of the exposure sequence. Under the Embedding & MLP paradigm, we map the time-series S^u to a sequence of numerical observations:

$$E^u = [e_1^u, e_2^u, \dots, e_N^u] \quad (2)$$

where $E^u \in \mathbb{R}^{d \times N}$, $e_i^u \in \mathbb{R}^d$, and e_i^u is the dense representation of the item at the i -th exposure with the fixed length d , mapped by the embedding layer. Then the goal of exposure sequence modeling is to design and learn a mapper Θ :

$$w^u = \Theta(E^u) \quad (3)$$

where $w^u \in \mathbb{R}^d$ is the embedding of user u representing her interest implied in exposure histories. Theoretically, Θ could be chosen or adapted from sequential modeling methods, and among which the mostly used are RNN and Transformer.

2.2 Exposure Sequence Modeling with Gama

Considering the noise and the long length of the exposure sequence, the mostly used RNN and Transformer methods bring noise and high latency problems. To address these problems, We propose Gama which adopts gating-adapted wavelet multiresolution Analysis (MRA) to model exposure sequence from a perspective of signal processing. In this section, **we reformulate the exposure sequence modeling in the framework of signal processing and then detail the proposed gating adaption.**

2.2.1 Exposure Signal Decomposition. The noise, i.e., the isolated exposures in the overall user exposure signals, is generally high frequency spikes. To reduce the noise, we first decompose user exposure signals into different frequency sub-bands. The illustration of Gama is shown in Figure 1. Gama decomposes user exposure sequence E^u into multiple components with different frequencies following the pyramidal algorithm framework [6]. Specifically, we suppose the resolution of the signal E^u to be 1, and the wavelets act as quadrature mirror filters:

$$a_k^1 = \sum_p \tilde{H}_p e_{p+2k}^u, \quad d_k^1 = \sum_p \tilde{G}_p e_{p+2k}^u \quad (4)$$

where \tilde{H} and \tilde{G} are constant vectors acting as the low-pass and high-pass filters, respectively. \tilde{H}_p stands for the p -th element in \tilde{H} .

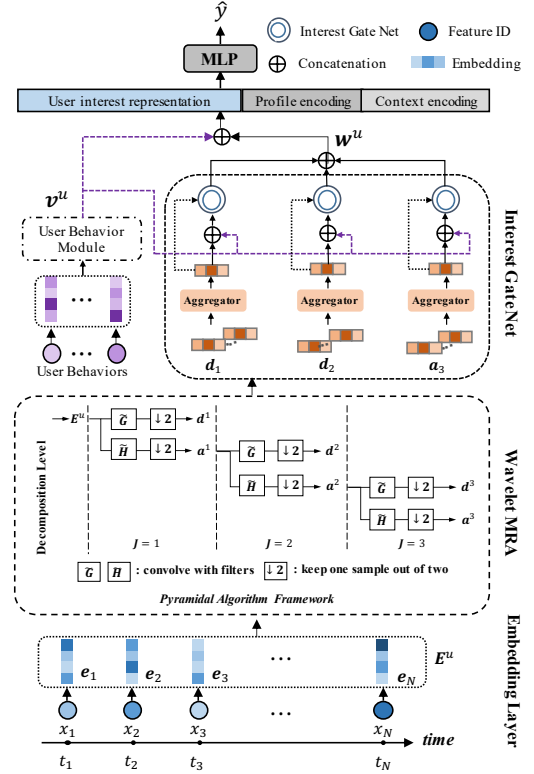


Figure 1: An illustration of the implementation of Gama in a CTR Prediction Model.

At the decomposition level $J = 1$, $a^1 \in \mathbb{R}^{d \times N/2}$ is the first low-frequency approximation with $1/2$ resolution, of the original signal E^u , and $d^1 \in \mathbb{R}^{d \times N/2}$ is the first high-frequency detail. $a_k^1 \in \mathbb{R}^d$ refers to the k -th sequential position in a^1 . Then the low-frequency approximation can be decomposed hierarchically to obtain recursive approximations and details at a series of lower resolutions:

$$a_k^n = \sum_p \tilde{H}_p a_{2k+p}^{n-1}, \quad d_k^n = \sum_p \tilde{G}_p a_{2k+p}^{n-1} \quad (5)$$

where $a^n \in \mathbb{R}^{d \times N/2^n}$ is the n -th approximation with resolution being $1/2^n$, and $d^n \in \mathbb{R}^{d \times N/2^n}$ is the n -th detail.

Choosing an appropriate wavelet base is important for separating the intrinsic features and patterns in exposure sequence. The quadrature mirror filters \tilde{G} and \tilde{H} both differ according to the wavelet base. Each wavelet base has its own non-stationary characteristic, and the convolution sum between the user exposure sequence signal and the wavelet base measures the similarity between the signal and the wavelet base [7]. Meanwhile, a suitable decomposition level can boost performance. According to the previously practical experience, too many decomposition layers would reduce the internal regularity of the signal, while too few can not effectively separate approximations and details. As shown in Figure 1, we take the decomposition level to be 3 as an illustration.

From the perspective of denoising, the highest-frequency detail d^3 is abandoned. Since a user might hardly pay attention to exposures with no connection with any aspect concerned by the user, these isolated exposures are disturbing high frequency spikes in the whole exposure sequence signal.

In this way, we can capture multi-dimension user interest by making full use of the multiresolution components decomposed from user exposure sequence, i.e., $\mathbf{a}^3, \mathbf{d}^1, \mathbf{d}^2$. The components with different frequencies characterize multiple intrinsic patterns of the origin user exposure sequence. With a high frequency, the components correspond to the transitory patterns of user interest, such as brand-level preferences. The steady patterns are implied in the components with a low frequency, such as style-level preferences.

2.2.2 Interest Gate Net. The multiresolution components, i.e., $\mathbf{a}^3, \mathbf{d}^1, \mathbf{d}^2$, are each aggregated along the axis *time*. The aggregator is originally average pooling (G-Ave), and can be adapted to a learnable attention module (G-Att) flexibly.

Let $\mathbf{v}^u \in \mathbb{R}^d$ denotes the embedding of user u representing her interest extracted from her behavior histories. To better extract user interest which is complementary to \mathbf{v}^u , We then design the Interest Gate Net module based on gating mechanism as shown in Figure 1. Has been motivated by the forget gate in LSTM [3], our proposed Interest Gate Net can further denoise and extract the most important user interest information from exposure signals by reweighting the multiple components with different frequencies, adaptive to user behavior sequence. The final user interest representation \mathbf{w}^u is obtained by concatenating the outputs of this Gate Net:

$$\mathbf{w}^u = [\mathcal{G}(\text{Att}(\mathbf{e}^q, \mathbf{d}^1), \mathbf{v}^u), \mathcal{G}(\text{Att}(\mathbf{e}^q, \mathbf{d}^2), \mathbf{v}^u), \mathcal{G}(\text{Att}(\mathbf{e}^q, \mathbf{a}^3), \mathbf{v}^u)] \quad (6)$$

Here the attentive aggregator $\text{Att}(\cdot)$ is taken for illustration, in which features of the target item \mathbf{e}^q are used as the query:

$$\text{Att}(\mathbf{e}^q, \mathbf{s}) = \sum_{t=1}^T \alpha_t \mathbf{s}_t, \quad \alpha_t = \frac{\exp(\mathbf{e}^q \mathbf{W} \mathbf{s}_t)}{\sum_{\tau=1}^T \exp(\mathbf{e}^q \mathbf{W} \mathbf{s}_\tau)} \quad (7)$$

where \mathbf{W} denotes the parameters of the attentive aggregator. The Interest Gate Net module $\mathcal{G}(\cdot)$ is designed to distill the potentially multi-dimension aggregated interests of users from exposure sequence information, which is concatenated with \mathbf{v}^u as inputs:

$$\mathcal{G}(\mathbf{s}, \mathbf{v}^u) = \sigma(\mathbf{W}_G^T [\mathbf{s}, \mathbf{v}^u] + \mathbf{b}_G) \odot \mathbf{s} \quad (8)$$

where σ is the logistic sigmoid function, \odot denotes the element-wise product, and \mathbf{W}_G^T is the weight matrix and \mathbf{b}_G is the bias vector for the Gate module $\mathcal{G}(\cdot)$.

It is easy to see the primary advantages of Gama. We capture the multi-dimension user interest from user exposure sequence and reduce the noise by decomposing user exposure sequence to multiresolution components. Besides, we extract the most important user interest information from exposure signals and further reduce the denoise using our proposed Interest Gate Net module. Moreover, all the exposure sequence modeling operations are completed with linear computational complexity.

2.2.3 Computational Complexity Analysis. Our proposed Gama achieves favorable performance when it comes to computational complexity. This is an important advantage since it is not feasible to apply traditional sequential modeling method such as RNN and Transformer to extremely dense user exposure sequence. As is well known, the time complexity of Transformer [9] is $O(dN^2)$, where N is the length of exposure sequence used and d is the fixed length of vector representation. The time complexity is $O(dLN)$ for Gama-Avg and $O(dL(N+1))$ for Gama-Att, where $L \ll N$ is the constant length of wavelet filters. Generally, the constant L equals to 2 for Haar wavelet, and 6 for Daubechies3.

Table 1: Statistics of Taobao and Real Production Datasets.

Taobao Dataset	#Users 925491	#Items 652648	#Samples 1.65M
	#Campaigns 423436	#Brands 99815	#Categories 6769
Real Production	#Users 62M	#Items 13M	#Samples 1.2 billions
	#Features 200	#Feature Vocabularies 0.1 billions	

3 EXPERIMENTS

3.1 Datasets and Experimental Settings

3.1.1 Public Dataset. Taobao Dataset¹ is a widely used benchmark dataset in CTR prediction works. Logs from 2017-05-09 to 2017-05-12 are for training and logs on 2017-05-13 are for testing. For performance evaluation, we generate two testing datasets: 1) All, 2) Cold. ‘‘All’’ is randomly sampled including 330 million samples. An user is ‘‘cold’’ when she has no click behavior, and the Cold users test datasets includes 0.1 million samples. For user profile feature, we use user ID, age and gender. For item profile, we use item ID, campaign ID, category ID and brand ID. The dataset statistics after preprocessing are shown in Table 1.

3.1.2 Real Production Dataset. The real production datasets for offline evaluations are traffic logs collected from our display advertising system. 1.2 billion exposure/click logs in the first 15 days are for training, and 0.5 million from the following day for testing. We use full feature set including user histories, user profiles, item profiles and contexts. User click sequences in previous 30 days are used for user behavior modeling, and user exposure sequence in the latest 3 days are for exposure sequence modeling. The dataset statistics after preprocessing are shown in Table 1.

3.1.3 Compared Algorithms. DNN[1], DIN[13], DIEN[12], and DSIN[2] are representatives in CTR prediction, without exposure modeling originally, thus they are used as baselines as well as backbones to further validate the adaptability and effectiveness of our Gama. Meanwhile, we also adopt DNN-NU the variant of DNN, as the baselines. DFN[11] is the only SOTA framework that utilizes different types of sequential user feedback such as exposure, click sequence, etc. For experiments on real production dataset, our online user interest model UIM is adopted as baseline, which is a DSIN-like model which has been highly optimized for our online system. **DNN-NU:** DNN without user historical click behavior sequence. This is used for verifying the contribution of Gama as the only user histories in the CTR prediction. **UIM:** UIM conducts multi-head self-attention on user click sessions. To meet the strict latency requirement, instead of the Bi-LSTM in DSIN, UIM performs vanilla attention across the whole user behavior histories to generate user interest representation.

3.1.4 Experimental Settings. For evaluations both on Taobao dataset and our real production dataset, and for the A/B testing in our real-world production environment as well, Daubechies3 is used as the wavelet base, and the decomposition level is set to 3. For evaluations on Taobao dataset, all algorithms adopt 128 as the length of exposure sequence while other parameters are followed by [2]. For experiments on the real production dataset, both \mathbf{w}^u and \mathbf{v}^u are 16-dimensional vectors, all the feature vectors are then feed into a 4-layer MLP with dimension of 1024, 512, 256, 1.

¹<https://tianchi.aliyun.com/dataset/dataDetail?dataId=56>

Table 2: Results when applying Gama to representative CTR prediction models on Taobao dataset.

BaseModel	Dataset	Origin	+AP-e	+Trans-e	+G-Avg★	+G-Avg	+G-Att★	+G-Att	AUC	RelaImpr
DNN-NU	All	0.6147	0.6184	0.6193	0.6232	0.6252	0.6302	0.6324 †	15.43%	
	Cold	0.5680	0.5737	0.5758	0.5789	0.5801	0.5879	0.5901 †	32.5%	
DNN	All	0.6221	0.6249	0.6262	0.6289	0.6299	0.6335	0.6347 †	10.31%	
	Cold	0.5755	0.5809	0.5833	0.5851	0.5863	0.5910	0.5929 †	23.04%	
DIN	All	0.6244	0.6271	0.6281	0.6306	0.6323	0.6350	0.6370 †	10.12%	
	Cold	0.5774	0.5830	0.5850	0.5872	0.5887	0.5920	0.5942 †	21.70%	
DIEN	All	0.6271	0.6301	0.6318	0.6338	0.6346	0.6371	0.6388 †	9.20%	
	Cold	0.5796	0.5853	0.5868	0.5894	0.5914	0.5959	0.5967 †	21.48%	
DSIN	All	0.6295	0.6321	0.6340	0.6353	0.6368	0.6403	0.6431 †	10.50%	
	Cold	0.5811	0.5861	0.5885	0.5910	0.5923	0.5941	0.5965 †	18.99%	

Note: † indicates that the improvement is statistical significant at the significance level of 0.05 over +Trans-e on AUC.

Table 3: Performance in DFN Architecture on Taobao dataset.

Model	Components in DFN Architecture		AUC
	Behavior Sequence	Exposure Sequence	
DFN	G-Att	-	0.6241
		-	0.6262
	Transformer	AP-e	0.6281
		DIN	0.6292
		DIEN	0.6291
		DSIN	0.6307
		Transformer	0.6310
		G-Avg	0.6344 †
		G-Att	0.6367 †

Note: † indicates that the improvement is statistical significant at the significance level of 0.05 over DFN.

3.1.5 Metrics. We adopt AUC and RelaImpr [13] as the evaluation metrics which are widely used in CTR prediction tasks. For a fair comparison, every model is repeatedly trained and tested 5 times and the average results are reported.

3.2 Experimental Results

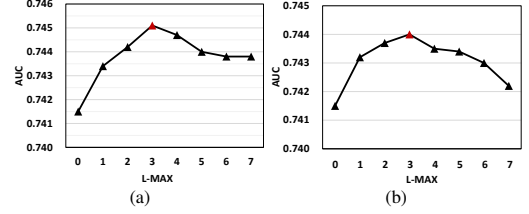
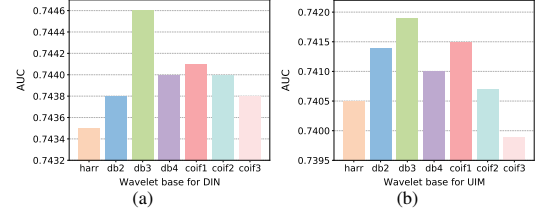
3.2.1 Effectiveness and Adaptability. To validate the effectiveness the adaptability of Gama to various networks, we plug Gama in many representative networks in CTR prediction. To further illustrate the effect of Interest Gate Net, we compare the performance of Gama and two variants of Gama. G-Avg★ and G-Att★ refer to Gama-Avg and Gama-Att without Interest Gate Net, respectively. According to Table 2, Gama could improve all of the BaseModels, and it brings at least 9.20% AUC improvement. Interest Gate Net can further improve the performance of Gama consistently. Meanwhile, there is additional performance gain brought by Gama for cold-start users. Gama can bring up to 18.99% performance improvement on “Cold-u” dataset. This is basically consistent with our analysis since for user who is not active, her exposure histories are valuable information to character her interest more accurately.

To validate the capability of Gama for exposure sequence modeling compared to the state-of-the-art, we replace the corresponding Transformer module in DFN architecture with Gama. The comparison results are shown in Table 3. With further analysis, we can find that Gama beats Transformer in exposure sequence modeling through de-noising and multi-dimension decomposition. However, when applied to model user click behavior sequence, Gama is weaker than Transformer. It implies that Gama is more competent for denser and noisier exposure sequence, which is one of our main motivation.

3.2.2 Experiments and Analysis on Industrial Dataset & Online Deployment. To exam the effectiveness of Gama in a real industrial CTR prediction scenario, we conduct both offline evaluation and online A/B testing. Table 4 shows the experimental results.

Table 4: Experiments in real-world production environment.

Offline	Origin	+AP-e	+Trans-e	+G-Avg	+G-Att
DIN	0.7340	0.7379	0.7390	0.7419 †	0.7439 †
DFN(click)	0.7348	0.7380	0.7391	0.7417 †	0.7431 †
UIM	0.7377	0.7410	0.7413	0.7446 †	0.7463 †
Online A/B Test		CTR gain		CPM gain	
UIM		-		-	
UIM + TRS-e		1.93%		1.79%	
UIM + G-Avg		2.97%		2.84%	
UIM + G-Att		3.67%		3.75%	

**Figure 2: Decomposition level effect on AUC.****Figure 3: Effect of Wavelet Base on AUC. Two different networks are used as BaseModels: (a) DIN and (b) UIM.**

Gama contributes 3.67% CTR and 3.75% CPM (Cost Per Mille) gain when plugged into UIM which has already been highly optimized for our online system with abundant of features. Note that 1% increase in CPM is critical for our full-fledged advertising platform, and the corresponding advertisement income of our advertising platform can be increased by 40 million dollars in one year.

3.2.3 Effect of Hyper-Parameters. We evaluate the effect of the decomposition level, wavelet base choosing and the length of exposure sequence. **Decomposition Level:** Haar and Daubechies3 are used as the wavelet bases. As shown in Figure 2, we can come to an empiricism that the decomposition level to 3 is acceptable for generally industrial scenarios. **Wavelet Base:** The decomposition level is set to be 3 and the length of user exposure sequence is 512. As shown in Figure 3, Experiments are conducted on both DIN and UIM. In these experiments, the most widely used wavelets are compared, such as Daubechies(db)2-4, Coiflet(coif)1-3, Harr(harr), among which db3 performs the best. A possible explanation for the higher effectiveness of db3 is that it’s non-stationary characteristic is more close to exposure signal. **Sequence Length:** According to Figure 4, for Gama and the BaseModels, i.e., AP-e and Trans-e, longer exposure sequence results in higher AUC. This is reasonable since longer exposure sequence is able to carry more information.

3.2.4 Empirical Analysis of Complexity. Besides the theoretical analysis on the high efficiency of Gama, we conduct experiments in our CPU-based online prediction system for empirical evaluations. TP99 latency is usually used for performance monitoring in commercial systems, which means the minimum time under which 99% of requests have been served. As shown in Figure 4, we plot the online TP99 latency of Gama in comparison with transformer and average

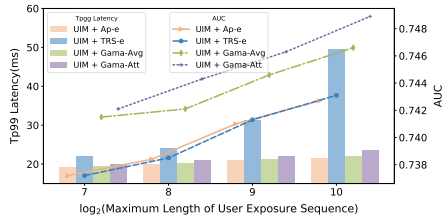


Figure 4: TP99 latency in real online CTR prediction system w.r.t length of exposure sequence.

pooling when plugged into UIM. Results in Figure 4 demonstrate that directly applying transformer to exposure sequence is not feasible for production environment when strict latency condition is required, which empirically verify that Gama is more suitable for user exposure histories modeling due to its much lower complexity.

4 CONCLUSIONS

We propose Gama to effectively address the high latency and noise problems in user exposure sequence modeling for CTR prediction. Specifically, Gama decomposes exposure sequence into multi-frequency components utilizing wavelet multiresolution analysis (MRA). Besides abandoning the highest frequency component, Gama adapts wavelet MRA with Interest Gate Net to further denoise and distill multi-dimension user interest adaptive to user behavior histories. Extensive experiments are conducted and the results demonstrate the low latency and effectiveness of our proposed method. In fact, Gama has been deployed in our real-world large-scale industrial advertising recommender and successfully serving over hundreds of millions of consumers for online e-commerce service. This is the first work to integrate non-parametric wavelet MRA method into deep neural networks to model user sequential exposure, and more adaptations in multi-modal[4] and heterogeneous graph-based[10] recommendation scenarios will be further explored.

REFERENCES

- [1] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of RecSys*. 191–198.
- [2] 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 28th International Joint Conference on Artificial Intelligence*. AAAI Press, 2301–2307.
- [3] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
- [4] Chenyi Lei, Yong Liu, Lingzi Zhang, Guoxin Wang, Haihong Tang, Houqiang Li, and Chunyan Miao. 2021. SEMI: A Sequential Multi-Modal Information Transfer Network for E-Commerce Micro-Video Recommendations. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 3161–3171.
- [5] Fuyu Lv, Mengxue Li, Tonglei Guo, Changlong Yu, Fei Sun, Taiwei Jin, and Keping Yang. 2020. Unclicked User Behaviors Enhanced Sequential Recommendation. *arXiv preprint arXiv:2010.12837* (2020).
- [6] S. G. Mallat. 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11, 7 (1989), 674–693.
- [7] Wai Keng Ngui, M Salman Leong, Lim Meng Hee, and Ahmed M Abdelrhman. 2013. Wavelet analysis: mother wavelet selection methods. In *Applied Mechanics and Materials*, Vol. 393. Trans Tech Publ, 953–958.
- [8] Pi Qi, Xiaoqiang Zhu, Guorui Zhou, Yujing Zhang, Zhe Wang, Lejian Ren, Ying Fan, and Kun Gai. 2020. Search-based User Interest Modeling with Lifelong Sequential Behavior Data for Click-Through Rate Prediction. In *Proceedings of the 2020 ACM Conference on Information and Knowledge Management*.
- [9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*. 5998–6008.
- [10] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous graph attention network. In *The World Wide Web Conference*. 2022–2032.
- [11] Ruobing Xie, Cheng Ling, Yalong Wang, Rui Wang, Feng Xia, and Leyu Lin. 2020. Deep Feedback Network for Recommendation. *Proceedings of IJCAI-PRICAI* (2020).
- [12] 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 *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 5941–5948.
- [13] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. 2018. Deep interest network for click-through rate prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1059–1068.