Deep Context Interest Network for Click-Through Rate Prediction

Xuyang Hou,Zhe Wang,Qi Liu,Tan Qu,Jia Cheng,Jun Lei

Meituan, Beijing, China

{houxuyang,wangzhe65,liuqi104,qutan,jia.cheng.sh,leijun}@meituan.com

ABSTRACT

Click-Through Rate (CTR) prediction, estimating the probability of a user clicking on an item, is essential in industrial applications, such as online advertising. Many works focus on user behavior modeling to improve CTR prediction performance. However, most of those methods only model users' positive interests from users' click items while ignoring the context information, which is the display items around the clicks, resulting in inferior performance. In this paper, we highlight the importance of context information on user behavior modeling and propose a novel model named **D**eep **C**ontext **I**nterest **N**etwork (DCIN), which integrally models the click and its display context to learn users' context-aware interests. DCIN consists of three key modules: 1) Position-aware Context Aggregation Module (PCAM), which performs aggregation of display items with an attention mechanism; 2) Feedback-Context Fusion Module (FCFM), which fuses the representation of clicks and display contexts through non-linear feature interaction; 3) Interest Matching Module (IMM), which activates interests related with the target item. Moreover, we provide our hands-on solution to implement our DCIN model on large-scale industrial systems. The significant improvements in both offline and online evaluations demonstrate the superiority of our proposed DCIN method. Notably, DCIN has been deployed on our online advertising system serving the main traffic, which brings 1.5% CTR and 1.5% RPM lift.

CCS CONCEPTS

• **Information systems** → **Personalization**.

KEYWORDS

CTR Prediction, User Behavior Modeling, Context-aware Interests

1 INTRODUCTION

Click-Through Rate (CTR) prediction, which aims at predicting the probability of a user clicking on an item, is critical for industrial
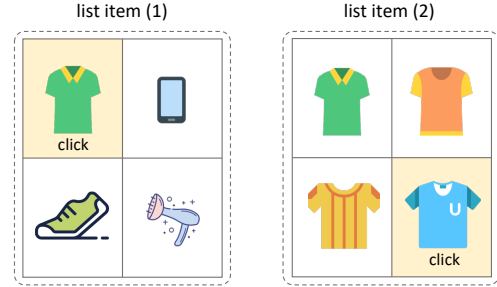
Figure 1: The illustration of display context. User's click behavior on an item may be influenced by other display items.

applications such as online advertising and recommendation systems [5, 16]. Recent research on CTR can be divided into many directions, such as feature interaction and user behavior modeling. This paper focuses on the problem of user behavior modeling.

Works on user behavior modeling hypothesize that users' historical behaviors reflect users' interests towards items. Lots of works [7, 11, 13, 18, 21, 22] focus on capturing user interests from click behaviors and achieve great success in CTR prediction. However, some studies [13, 15] find that there are gaps between users' click behaviors and their real interests, leading to biased user representation and inferior performance. In order to tackle the problem, pioneering works DFN [17], DUMN [1] introduce users' unclick behaviors to capture users' negative preferences and get further performance promotion. However, they model users' click and unclick behaviors separately, which ignores the interaction between click and unclick items. As shown in Figure 1, there are generally many display items on one screen of a computer/smartphone. Users' click behaviors are not only determined by the matching degree between items and their interests but also influenced by the display context [2, 12, 20]. For example, as shown in Figure 1, when a user wants to buy a T-shirt, he may click the green one when its surrounding items are smartphones, shoes, or hair dryers since those are boring to him. But when the display items are all T-shirts, he may click the blue one because it may match his interests better. Based on the above analysis, we argue that display context is also essential in interest modeling. RACP [6] introduces page-wise feedback modeling and aggregates different pages into different interest representations. However, different clicks on one page may indicate different interests, which degrades the model's capability to capture users' fine-grained interests precisely.

In this paper, we propose **D**eep **C**ontext **I**nterest **N**etwork (DCIN) to extract context-aware interest representations from users' click behaviors and corresponding display contexts aiming at acquiring comprehensive and fine-grained user representation. Specifically, we first design a Position-aware Context Aggregation Module

(PCAM), which performs aggregation of display items with an attention mechanism. Meanwhile, we consider position bias [9, 10] and integrate position information into PCAM. Then, Feedback-Context Fusion Module (FCFM) is proposed to fuse the representation of clicks and corresponding display contexts through non-linear feature interaction to capture users' context-wise interests. Finally, we use Interest Matching Module (IMM) to match behaviors related to the target item. Moreover, we also provide our hands-on solution to implement DCIN on large-scale industrial systems. Our contributions are as follows:

- We highlight the necessity of introducing context information for interest modeling, which can be used to capture users' interests more precisely and comprehensively.
- We devise a novel sequential model, DCIN, which effectively integrates display context and position information to acquire users' context-aware interests.
- We conduct both offline and online experiments, and the significant improvements demonstrate the superiority of our proposed DCIN method. The online optimization makes DCIN process users' behavior sequences 28x longer than the SOTA method RACP [6]. Notably, DCIN has served the main traffic in our online advertising system, bringing 1.5% CTR and 1.5% RPM lift.

2 METHODOLOGY

In this section, we introduce the design of DCIN.

2.1 Preliminaries

CTR prediction is to predict the probability p that user u clicks target item t with model $\mathcal{F}$, given user behaviors C (click items) and D (display items), and context c, formulated as $p = \mathcal{F}(C, D, u, t, c)$. The click behaviors are a sequential list of click items with corresponding features such as item id, category, etc., i.e., $C = \{c_1, c_2 \ldots, c_N\}$, where c_i is the i-th click item and N is the length of C. To model the display context, we pick out M corresponding display items for each click, hence display behaviors D can be denoted as:

$$D = \{[d_{11}, d_{12}, \ldots, d_{1M}] \ldots, [d_{N1}, d_{N2}, \ldots, d_{NM}]\}, \quad (1)$$

where d_{ij} is the j-th display item around click c_i. We try to keep the click in the middle when selecting display items for each click. We denote p_i and q_{ij} as the absolute position of c_i and d_{ij} respectively. We also calculate relative position $r_{ij} = p_i - q_{ij}$ between c_i and d_{ij}. We utilize the widely-used embedding technique to transform sparse features into low-dimensional dense vectors. For example, e_{c_i} represents the embedding of click item c_i.

2.2 Deep Context Interest Network

The architecture of DCIN is shown in Figure 2, mainly consisting of three modules. Specifically, we first design Position-aware Context Aggregation Module (PCAM) to aggregate the display items with an attention mechanism. Then Feedback-Context Fusion Module (FCFM) is proposed to fuse the representation of clicks and corresponding display contexts through non-linear feature interaction to acquire context-aware interests. Finally, we propose the Interest Matching Module (IMM) to learn the interests related to the target item adaptively.

2.2.1 ***Position-aware Context Aggregation Module***. As shown in Figure 1, there are generally many display items on one screen of a computer/smartphone. Users' click behaviors are not only determined by the matching degree between items and their interests but also influenced by the display context [2, 20]. However, modeling users' interests only from their click behaviors ignores the context information, leading to biased interest representation and inferior performance. To capture users' context-aware interests, we design PCAM, which aggregates display items for each click with an attention mechanism. Moreover, users generally tend to click top items on the screen (known as position bias [9, 10]) regardless of the relevance. Hence, items' positions also impact users' click behaviors. We also integrate position information into PCAM for a better understanding of users' actual interests. As a result, PCAM can be formulated as follows:

$$v_i = \sum_{j=1}^{M} \mu_{ij} W_V e_{d_{ij}}, \quad (2)$$

where v_i represents the display context corresponding to click c_i and is modeled as a weighted sum aggregation of display items $e_{d_{ij}}$, and W_V is learning parameters. μ_{ij} is the attention weight, which can be formulated as follows:

$$\mu_{ij} = \frac{exp(\alpha_{ij})}{\sum_{j=1}^{M} exp(\alpha_{ij})}, \quad (3)$$

$$\alpha_{ij} = W_Q x_{c_i} \cdot W_K x_{d_{ij}} + e_{r_{ij}}, \quad (4)$$

where α_{ij} is the relevance between the i-th click and its j-th display item, $x_{c_i} = concat(e_{c_i}, e_{p_i})$ represents the concatenation of item embedding e_{c_i} and its absolute position embedding e_{p_i} and the same as $x_{d_{ij}}$, $e_{r_{ij}}$ is the one-dimensional embedding vector of relative position p_{ij}, and W_Q and W_K are learning parameters.

2.2.2 ***Feedback-Context Fusion Module***. Previous user behavior models [7, 11, 13, 18, 21, 22] directly regard users' click items as users' interest, ignoring the context information. To capture users' context-aware interests, we propose FCFM to fuse the representation of clicks and corresponding display contexts through non-linear feature interaction. Concretely, FCFM is implemented by a 2-layer MLP:

$$f_i^{(k+1)} = ReLU(W^{(k)} f^{(k)} + b^{(k)}), \quad (5)$$

where $f_i^{(k)}$ is the output of the k-th layer, and $W^{(k)}$ and $b^{(k)}$ are learning parameters. The input is formulated as follows:

$$f_i^{(0)} = concat(x_{c_i}, v_i, x_{c_i} - v_i, x_{c_i} * v_i), \quad (6)$$

where $*$ represents the element-wise product. Then, we obtain context-aware interest $I_i = f_i^2$ from click c_i and its corresponding display context v_i. For generalization and trainability, we share the learnable parameters across different clicks.

2.2.3 ***Interest Matching Module***. With the first two modules, we integrally model the click and its display context to capture users' context-aware interests, denoted as $I = \{I_1, I_2, \ldots, I_N\}$, which are more precise and comprehensive than those extracted only from click behaviors. To capture the user interests that are related to the target item, the SOTA sequential models can be easily integrated, like DIN [22], DIEN [21], BST [3], etc. Here we choose the DIN
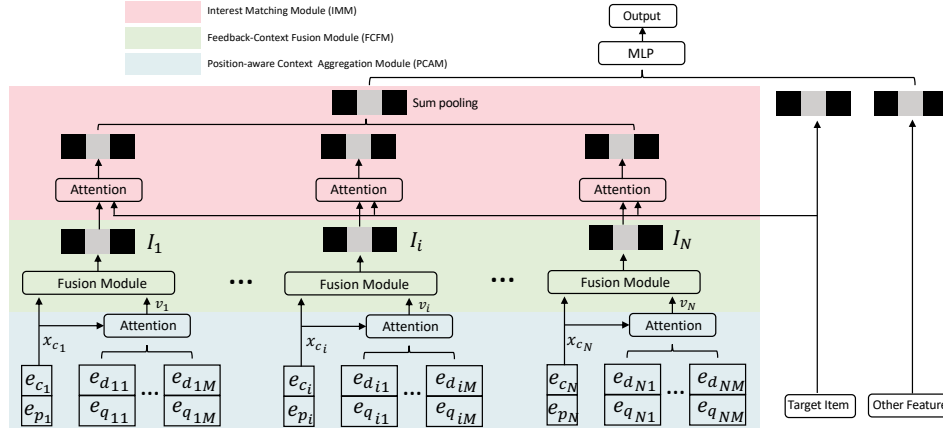
Figure 2: The architecture of DCIN network.

method to aggregate user interests, which can be formulated as follows:

$$U = \sum_{i=1}^{N} \frac{exp(\boldsymbol{e}_t \cdot \boldsymbol{I}_i)}{\sum_{j=1}^{N} exp(\boldsymbol{e}_t \cdot \boldsymbol{I}_j)} \cdot \boldsymbol{I}_i, \tag{7}$$

where U is the final user representation, which will be concatenated with other features and fed into MLP for CTR prediction. We utilize the widely-used negative log-likelihood function as the loss function to train our DCIN model.

2.3 Optimization for Online Serving

For industrial recommender systems, the serving latency constraints are highly restricted. Since we introduce display items to capture users' context-aware interests, the sequence length has been extended tens of times. Generally, the serving latency of online models is proportional to the sequence length. As a result, it is challenging to make DCIN serve in an industrial online system directly. To address the challenge, we make some online optimizations. Specifically, since PCAM and FCFM have no connection with the target item, we can pre-calculate users' context-aware interest representations $I = \{I_1, I_2, \ldots, I_N\}$, and directly make it as model's input when online serving, which saves a large amount of online serving time (10 milliseconds). Since RACP [6] needs the target item to aggregate page-wise interests, it can only shorten the sequence length to make the model serve online. With online optimization, DCIN can process users' behavior sequences 28x longer than the SOTA method RACP, adding only one millisecond of online latency.

3 EXPERIMENTS

3.1 Experimental Settings

3.1.1 Datasets. Integrating display context and position information into interest models is a novel exploration. Hence we build a 31-day industrial dataset from our online advertising system, which contains more than 1 billion samples. Following the standard industry practice, we split the data in the first 30 days as the training set and the data on the last day as the testing set. Users' recent 50-click

behaviors are reserved. The surrounding 20 display items for each click are utilized to model the display context.

3.1.2 Compared Methods. We implement six state-of-art models to compare with DCIN. All models use the same features for fair comparisons.

Wide&Deep [4] consists of a Wide part and a Deep part to learn raw and high-order feature interactions.

DeepFM [8] replaces the Wide part in [4] with an FM layer.

DIN [22] introduces an attention mechanism to capture users' diverse interests for different target items.

DIEN [21] proposes an interest evolving layer based on a gated recurrent unit (GRU) to model users' dynamic interests.

DFN [17] tries to learn positive preferences from click behaviors and negative preferences from unclick behaviors.

RACP [6] introduces page-wise feedback modeling and is the published SOTA solution for user behavior modeling.

3.1.3 Evaluation Metrics. We utilize a widely-adopted metric Area Under Curve (AUC) for evaluation in the CTR prediction task. We further introduce RelaImpr [19] to measure the relative improvements over the base model:

$$RelaImpr = \frac{AUC(measured\ model) - 0.5}{AUC(base\ model) - 0.5} - 1 \tag{8}$$

3.2 Experimental Results

We repeat all experiments 5 times and report their average results. In experiments, the embedding dimension is 16 for each feature. The hidden sizes of MLP in FCFM and the output layer are 64/16, and 1024/512/128 respectively. We utilize Adam with the learning rate to be 1e-4, and the batch size is 3000.

3.2.1 Offline Performance Comparison. The experimental results are shown in Table 1, from which we can find that our proposed DCIN significantly outperforms all methods on AUC and achieves 21.24% RelaImpr over the base model. As introduced in Section 2.3, DCIN can process users' behavior sequences 28x longer than RACP. For a fair comparison, DCIN (short) uses the same sequence length as RACP, and DCIN uses a longer sequence as implemented in

our online advertising system. Specifically, first, DIN and DIEN perform much better than Wide&Deep and DeepFM because they capture users' diverse and temporal interests for different target items. Second, DFN beats DIN and DIEN slightly since it tries to capture dislike signals from unclick behaviors. Fourth, RACP performs better than DFN since it models users' contextualized page-wise interests. Fifth, our proposed DCIN (short) beats RACP. On the one hand, different clicks, even on one page, may indicate different interests. RACP only models page-wise interests, degrading the model's capability to capture users' fine-grained interests. On the other hand, RACP ignores the impact that items' positions have on users' click behaviors. Finally, DCIN beats DCIN (short) since it introduces a longer sequence.

**Table 1: Performance comparison with different methods.**

Model	AUC	RelaImpr
Wide&Deep	0.63588	0.00%
DeepFM	0.63602	0.10%
DIN	0.65012	10.48%
DIEN	0.65102	11.14%
DFN	0.65220	12.01%
RACP	0.65872	16.81%
DCIN (short)	0.66186	19.12%
DCIN*	**0.66475**	**21.24%**
DCIN (short)-w/o position	0.65850	16.64%
DCIN (short)-w/o FCFM	0.65950	17.38%

3.2.2 ***Ablation Tests***. We also conduct ablation tests to verify the effectiveness and necessity of different components in DCIN. Firstly, we remove the position information in PCAM and conduct experiment DCIN (short)-w/o position. As shown in Table 1, DCIN (short)-w/o position performs worse than DCIN (short) since items' positions also have important impact on users' click behaviors. Secondly, we remove the FCFM module in DCIN and conduct experiment DCIN (short)-w/o FCFM where we directly regard display context v_i as users' interests, which can be formulated as $I_i = v_i$ instead of $I_i = f_i^2$. As shown in Table 1, DCIN (short)-w/o FCFM also performs poorly than DCIN (short) due to the lack of fusion between clicks and display contexts.

3.2.3 ***Online A/B Test***. We conduct a strict online A/B test to evaluate DCIN on our online advertising system for two weeks. Compared with DIN (online baseline), DCIN has improved CTR by 1.5% and RPM (Revenue Per Mile) by 1.5%. For now, DCIN has been deployed online and serves the main traffic, which contributes to significant business revenue growth.

3.2.4 ***Case Study***. To demonstrate the effectiveness of DCIN for context-aware interest representations, we randomly select 10 click items and choose 100 different display contexts for each click respectively. Then we extract the corresponding context-aware interests I_i and visualize their distribution through t-SNE [14]. As shown in Figure 3, the distinct separation among click items indicates the

dominance of click behavior in learning. The obvious diversity of different contexts of the same click demonstrates the ability of DCIN to capture fine-grained context-aware interest representations.
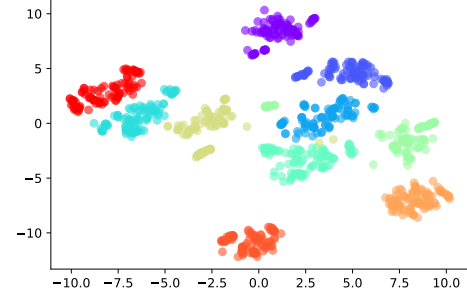
Figure 3: Visualization of context-aware interests for 10 click items under 100 different display contexts. Each color represents one click item. The representation distribution shows the separation among click items and the diversity among different display contexts of the same click item.

To verify the effectiveness of context modeling, we visualize the attention weights from IMM between the target item and the same click item under different display contexts. As shown in Figure 4, the weights of DIN (green color) is the same since DIN only takes click behaviors as input, resulting in the inability to distinguish different display contexts. While DCIN learns higher weights under the right competitive display context where all display items are of interest to the user.
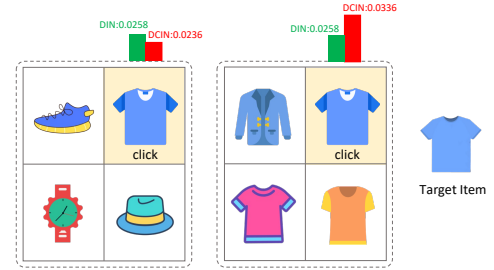
Figure 4: Case study of attention weights between the target item and the same click item under different display context. DIN produces the same attention weights due to missing context information. The proposed DCIN can assign higher weight to the click under competitive display context.

4 CONCLUSIONS

In this paper, we highlight the necessity of display contexts and position information for interest modeling and devise a novel sequential model DCIN. The significant improvements in both offline and online evaluations demonstrate the superiority of our proposed DCIN methods. Notably, DCIN has been deployed on our online advertising system, and now it is serving the main traffic.

REFERENCES

[1] Zhi Bian, Shaojun Zhou, Hao Fu, Qihong Yang, Zhenqi Sun, Junjie Tang, Guiquan Liu, Kaikui Liu, and Xiaolong Li. 2021. Denoising User-aware Memory Network for Recommendation. In *Fifteenth ACM Conference on Recommender Systems*. 400–410.

[2] Jesús Bobadilla, Raúl Lara-Cabrera, Ángel González-Prieto, and Fernando Ortega. 2020. Deepfair: deep learning for improving fairness in recommender systems. *arXiv preprint arXiv:2006.05255* (2020).

[3] Qiwei Chen, Huan Zhao, Wei Li, Pipei Huang, and Wenwu Ou. 2019. Behavior sequence transformer for e-commerce recommendation in alibaba. In *Proceedings of the 1st International Workshop on Deep Learning Practice for High-Dimensional Sparse Data*. 1–4.

[4] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. 2016. Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems*. 7–10.

[5] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems*. 191–198.

[6] Zhifang Fan, Dan Ou, Yulong Gu, Bairan Fu, Xiang Li, Wentian Bao, Xin-Yu Dai, Xiaoyi Zeng, Tao Zhuang, and Qingwen Liu. 2022. Modeling Users' Contextualized Page-wise Feedback for Click-Through Rate Prediction in E-commerce Search. (2022).

[7] 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. *arXiv preprint arXiv:1905.06482* (2019).

[8] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: a factorization-machine based neural network for CTR prediction. *arXiv preprint arXiv:1703.04247* (2017).

[9] Huifeng Guo, Jinkai Yu, Qing Liu, Ruiming Tang, and Yuzhou Zhang. 2019. PAL: a position-bias aware learning framework for CTR prediction in live recommender systems. In *Proceedings of the 13th ACM Conference on Recommender Systems*. 452–456.

[10] Jianqiang Huang, Ke Hu, Qingtao Tang, Mingjian Chen, Yi Qi, Jia Cheng, and Jun Lei. 2021. Deep Position-wise Interaction Network for CTR Prediction. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1885–1889.

[11] Jing Li, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Tao Lian, and Jun Ma. 2017. Neural attentive session-based recommendation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 1419–1428.

[12] Xiang Li, Shuwei Chen, Jian Dong, Jin Zhang, Yongkang Wang, Xingxing Wang, and Dong Wang. 2023. Decision-Making Context Interaction Network for Click-Through Rate Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence* 37, 4 (Jun. 2023), 5195–5202.

[13] Qi Pi, Weijie Bian, Guorui Zhou, Xiaoqiang Zhu, and Kun Gai. 2019. Practice on long sequential user behavior modeling for click-through rate prediction. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2671–2679.

[14] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).

[15] Menghan Wang, Mingming Gong, Xiaolin Zheng, and Kun Zhang. 2018. Modeling dynamic missingness of implicit feedback for recommendation. *Advances in neural information processing systems* 31 (2018).

[16] Zhibo Xiao, Luwei Yang, Wen Jiang, Yi Wei, Yi Hu, and Hao Wang. 2020. Deep multi-interest network for click-through rate prediction. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2265–2268.

[17] Ruobing Xie, Cheng Ling, Yalong Wang, Rui Wang, Feng Xia, and Leyu Lin. 2021. Deep feedback network for recommendation. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*. 2519–2525.

[18] Weinan Xu, Hengxu He, Minshi Tan, Yunming Li, Jun Lang, and Dongbai Guo. 2020. Deep interest with hierarchical attention network for click-through rate prediction. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1905–1908.

[19] Ling Yan, Wu-Jun Li, Gui-Rong Xue, and Dingyi Han. 2014. Coupled group lasso for web-scale ctr prediction in display advertising. In *International Conference on Machine Learning*. PMLR, 802–810.

[20] Yu Zheng, Chen Gao, Xiang Li, Xiangnan He, Yong Li, and Depeng Jin. 2021. Disentangling user interest and conformity for recommendation with causal embedding. In *Proceedings of the Web Conference 2021*. 2980–2991.

[21] 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.

[22] 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.