Deep Intention-Aware Network for Click-Through Rate Prediction

Yaxian Xia
yaxian.xyx@alibaba-inc.com
Alibaba Group

Yi Cao
dylan.cy@alibaba-inc.com
Alibaba Group

Sihao Hu*
sihaohu@gatech.edu
Georgia Institute of Technology

Tong Liu
yingmu@taobao.com
Alibaba Group

Lingling Lu
lulingling@zju.edu.cn
Zhejiang University

ABSTRACT

E-commerce platforms provide entrances for customers to enter mini-apps that can meet their specific shopping requirements. Trigger items displayed on entrance icons can attract more entering. However, conventional Click-Through-Rate (CTR) prediction models, which ignore user instant interest in trigger item, fail to be applied to the new recommendation scenario dubbed Trigger-Induced Recommendation in Mini-Apps (TIRA). Moreover, due to the high stickiness of customers to mini-apps, we argue that existing trigger-based methods that over-emphasize the importance of trigger items, are undesired for TIRA, since a large portion of customer entries are because of their routine shopping habits instead of triggers. We identify that the key to TIRA is to extract customers' personalized entering intention and weigh the impact of triggers based on this intention. To achieve this goal, we convert CTR prediction for TIRA into a separate estimation form, and present Deep Intention-Aware Network (DIAN) with three key elements: 1) *Intent Net* that estimates user's entering intention, *i.e.*, whether he/she is affected by the trigger or by the habits; 2) *Trigger-Aware Net* and 3) *Trigger-Free Net* that estimate CTRs given user's intention is to the trigger-item and the mini-app respectively. Following a joint learning way, DIAN can both accurately predict user intention and dynamically balance the results of trigger-free and trigger-based recommendations based on the estimated intention. Experiments show that DIAN advances state-of-the-art performance in a large real-world dataset, and brings a **9.39**% lift of online Item Page View and **4.74**% CTR for Juhuasuan, a famous mini-app of Taobao.

1 INTRODUCTION

Mini-app, a new marketing tool that can meet specific shopping requirements for customers, has been widely adopted in e-commerce platforms and contributes to billions of daily page views. Take Taobao for instance, Juhuasuan is a mini-app that only sells brand-discounted products, and Little Black Box is another mini-app that sells new products only. Another mini-app example is the Flash Deals of Shoppee, where customers can snap up low-priced products within time limits.

For e-commerce Apps, fixed entrances are always provided for customers to click and enter mini-apps. To further attract their attention, one or two trigger items are displayed on each entrance icon, which are personally recommended based on a user's historical behaviors, as shown in Figure 1(a). Once a user clicks the entrance icon, he/she will enter a mini-app feeds page where the
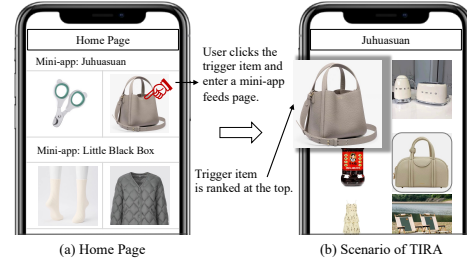

(a) Home Page   (b) Scenario of TIRA

Figure 1: Trigger-Induced Recommendation in Mini-Apps

clicked trigger item is ranked at the top, and other recommended items follow up, as shown in Figure 1(b). In this work, we focus on recommending follow-up items in the mini-app (feeds), given that a user has already clicked trigger item and entered a mini-app. We define this recommendation task as Trigger-Induced Recommendation in mini-Apps (TIRA).

The existence of trigger-item invalidates many popular CTR models like DIN[13], DIEN[12], SIM[6] and GIFT[1], because they cannot take into consideration the effect of trigger item that can reveal the instant interest when a user enters the mini-app.

Moreover, the characteristic of mini-app prevents trigger-based methods[5, 7, 9, 10] being applied to TIRA: Regular customers of mini-apps exhibit a strong stickiness to specific mini-apps, indicating that their entering behaviors are not likely affected by triggers but by their routine shopping habits. This observation can be proved by Figure 2: users who visit a mini-app more than 15 times a month have almost the same preference for the items from the trigger category and other categories. In this case, trigger-based methods that over-emphasize the importance of triggers and thus recommend follow-up items with high relevance to triggers are undesired. For regular users, recommendation results should exhibit diversity to meet their multiple interests to the mini-app.

Therefore, the most important objective for TIRA is to strike a balance between trigger-free and trigger-based recommendations via identifying customers' personalized entering intention. To achieve this goal, we present Deep Intention-Aware Network (DIAN) that explicitly recognizes user entering intention by an auxiliary task, and dynamically adjusts recommendation strategy. The whole DIAN consists of three key elements, dubbed Intention Net, Trigger-Aware Net and Trigger-Free Net. The Intention Net is designed to estimate a user's entering intention based on the his/her historical preference toward the trigger item or the mini-app. To improve the accuracy of intention score estimated by the Intention

---

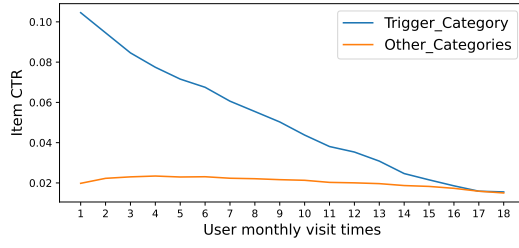*Sihao Hu is the corresponding author.

Figure 2: Average CTR of items from trigger category and other categories *w.r.t.* user monthly visit times in Juhuasuan.

Net, we introduce a posterior labeling strategy and an auxiliary training task, making our estimation can achieve 91.3% accuracy.

Trigger-Aware Net is proposed to estimate a conditional CTR by assuming that a user is attracted by the trigger item only. We elaborately design the model architecture to make it fully exploit the user's instant interest revealed by the trigger item. Conversely, Trigger-Free Net is proposed to estimate a conditional CTR by assuming a user's entering is only led by his/her routine habits and preference to the mini-app. Following a joint-learning paradigm, DIAN fuses the outputs of Trigger-Aware Net and Trigger-Free Net based on the estimated user intention score, making it dynamically adjust recommendation result according to customers with different entering intention.

Offline experiments show that DIAN advances state-of-the-art performance in a large real-world dataset. Online A/B testing suggests that DIAN can bring a **9.39**% lift of online Item Page View and **4.74**% CTR for Juhuasuan, a famous mini-app of Taobao. Currently, DIAN has been deployed on Taobao App since Sep. 21, 2022, serving for the full volume traffic of Juhuasuan.

Contributions. To summarize, the contributions are as follows:

- We define a new task of Trigger-Induced Recommendation in Mini-Apps (TIRA) and identify problems of existing trigger-free and trigger-based CTR prediction models for TIRA.
- We propose DIAN equipped with novel architecture and effective intention estimation strategy, making it can accurately reveal user's personalized intention and dynamically adjust the recommendation result.
- DIAN significantly advances the state-of-the-art on an large industrial dataset. Online A/B testing indicates DIAN can bring huge benefit to the real-world application.

2 PROPOSED METHOD

2.1 TIRA Problem Formulation

CTR prediction task aims to estimate a click-through probability $P(Click|Target)$, given a user be recommended a target item[3, 4, 11]. For TIRA, we take into consideration the impact of users' personalized entering intention to CTR estimation. We use $Intent = 0$ and $Intent = 1$ to indicate whether a user's entering intention is the mini-app or the trigger-item. Therefore, user intention can be represented as a probability distribution $(P(Intent = 0), P(Intent = 1))$, and $P(Intent = 0) + P(Intent = 1) = 1$ since there are only two reasons of entering the mini-app.

In this case, CTR estimation can be converted to another form by separately estimating 1) User intention, 2) CTR given user's intention to the trigger item, and 3) CTR given user's intention to

the mini-app:

$$P(Click \mid Target) = P(Click \mid Target, Intent = 0) * P(Intent = 0)$$
$$+ P(Click \mid Target, Intent = 1) * P(Intent = 1) \quad (1)$$

The advantage of this separate estimation way is that we can design individual neural networks to model user's interests to the target item given two different intentions, and leverage the estimated user intention score to balance two different types of user interests. Obviously, existing trigger-free and trigger-based methods cannot take both intentions into account.

To model above-mentioned three elements, we propose Deep Intention-Aware Network (DIAN), which is equipped with three key sub-nets: 1) Intention Net that aims to estimate user entering intention; 2) Trigger-Aware Net that aims to estimate CTR given user intention to the trigger-item; 3) Trigger-Free Net that aims to estimate CTR given user intention to the mini-app. We introduce them in detail one-by-one.

2.2 Intention Net

We start with introducing the strategy of identifying user entering intention in the training set.

User Intention Labeling: After entering the mini-app page, if a user still clicks on the trigger item or items with the same leaf-category (a relative strict category definition in Taobao), we label his/her entering intention to the trigger item ($Intent = 1$). Otherwise, we label the entering intention to the mini app ($Intent = 0$), because the user doesn't interested in the trigger item or similar items. It should be noted that although we can identify user entering intention, this posterior labeling strategy cannot be use for online serving because the recommendation happens at the moment the user enters the mini-app. Therefore, we need to estimate the intention score by Intention Net.

User Intention Estimation: To estimate the personalized intention score for each user, Intention Net takes four parts as the input: 1) User profile features like user_id, age, occupation, *etc.* 2) User short-term click item sequence, which reflects his/her historical interest. 3) User-mini-app cross features that measure the user's preference to the mini-app, including his/her monthly visit times or average stay duration. 4) Trigger item features like item_id, category_id, *etc.* We use widely adopted embedding technique to project them into embedding vectors as H_{user}, H_{sseq}, H_{crs}, and H_{tri}, respectively. The sequence vector H_{sseq} is the average pooling result of embeddings of items in the click sequence.

We concatenate four embedding vectors together and pass them through a MLP layer, which can be formalized as:

$$\hat{y}_{Intent} = \text{Sigmoid}\left(\text{MLP}(H_{user} \oplus H_{sseq} \oplus H_{crs} \oplus H_{tri})\right) \quad (2)$$

where $\hat{y}$ represents the user's intention probability $P(intent)$. $\oplus$ is the concatenate operator.

2.3 Trigger-Aware Net

Trigger-Aware Net (TAN) is proposed to estimate the CTR of the target item by assuming a user's entry is due to the trigger item. In this case, the trigger item is critical for recommending follow-up items. Therefore, we design TAN to highlight the impact of triggers and to fully exploit the instant interest revealed by triggers.
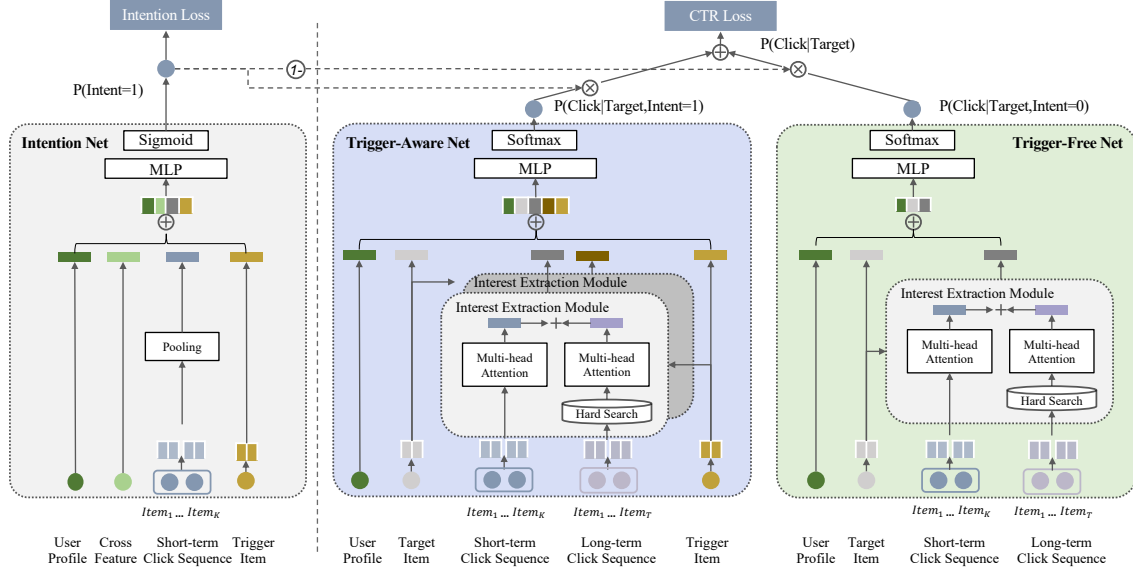
Figure 3: DIAN consists of three sub-nets: Intention Net, Trigger-Aware Net and Trigger-Free Net.

The input of TAN consists of five parts: 1) User profile features. 2) User short-term click sequence. 3) User long-term click sequence. 4) Trigger item features. 5) Target item features. Similarly as in Intention Net, we adopt embedding techniques to project these original features into embedding vectors and denote them as E_{user}, E_{sseq}, E_{lseq}, E_{tri} and E_{tar}. $E_{sseq} = [E_{item}^{s1}, \cdots, E_{item}^{sK}]$ and $E_{lseq} = [E_{item}^{l1}, \cdots, E_{item}^{lT}]$.

Dual-layer Interest Extraction: We extract user interest information contained within short-term and long-term click sequences separately according to the trigger item and target item, to prevent them from disturbing each other when they are diverse. For trigger item and target item, we generate user interest representation h_{tri} and h_{tar} respectively as follows:

$$H_{tri} = \text{InterestExtract}_{\text{tri}}(E_{tri}, E_{sseq}, E_{lseq}) \quad (3)$$

$$H_{tar} = \text{InterestExtract}_{\text{tar}}(E_{tar}, E_{sseq}, E_{lseq}) \quad (4)$$

Take Eq. 3 as an example, we introduce one layer of Interest Extraction (IE) in detail. An IE layer consists of short-term IE and long-term IE modules, where short-term IE module tackles click sequence of the past 14 days and long-term IE tackles click sequence of the past 180 days.

For short-term IE, we apply multi-head[8] target attention to retrieve relevant information within the short-term click sequence with the trigger item embedding E_{tri}, to generate a user short-term interest embedding H_{tri}^s:

$$head_i = W_i^V E_{sseq} \cdot \text{Softmax}\left(W_i^Q E_{tri} \odot W_i^K \cdot E_{sseq}\right)$$
$$H_{tri}^s = [head_1 \oplus head_2 \cdots \oplus head_n] \quad (5)$$

where $\odot$ is broadcasting dot-product operator, projection matrices for each head W_i^Q, $W_i^K \in \mathbb{R}^{d \times d/n}$ are learnable parameters and d is the dimension of hidden space.

User's long-term interest can bring more diverse recommendation results compared to short-term interest. To generate the long-term interest representation h_{tri}^l toward the trigger item, we

adopt the search-based IE method proposed by SIM[6]. Specifically, it hard-searches the most recent K items, S_{lseq}, belonging to the same leaf category of the trigger item, which can largely reduce the length of long-term click sequence. Similar to short-term IE, we apply another multi-head target attention function to softly retrieve relevant information within a filtered click sequence by the trigger item:

$$head_j = W_j^V S_{lseq} \cdot \text{Softmax}\left(W_j^Q E_{tri} \odot W_j^K \cdot S_{lseq}\right)$$
$$H_{tri}^I = [head_1 \oplus head_2 \cdots \oplus head_n] \quad (6)$$

It should be noted that projection matrices of long-term IE layer are not shared with short-term IE layer. The subscript $lseq$ and $sseq$ are omitted for the sake of simplicity. By adding short-term and long-term user interest embeddings, we obtain the user interest representation toward the trigger item, i.e., $H_{tri} = H_{tri}^{sseq} + H_{tri}^{lseq}$.

For the target item, just following the same calculation, we can obtain the user interest representation H_{tar}. Finally, we concatenate E_{user}, E_{tri}, E_{tar}, H_{tri} and H_{tar} together and pass them through a MLP layer:

$$\hat{y}_{tan} = \text{Sigmoid}\left(\text{MLP}(E_{user} \oplus E_{tri} \oplus E_{tar} \oplus H_{tri} \oplus H_{tar})\right) \quad (7)$$

where $\hat{y}_{tan}$ represents the CTR to target item, given user's entering intention is to trigger-item, i.e., $P(Click \mid Target, Intent = 1)$.

2.4 Trigger-Free Net

Trigger-Free Net (TFN) aims to estimate CTR given user's intention is to the mini-app. As shown in Figure 3, we adopt a similar architecture of TAN but exclude trigger item features from the input and remove the interest extraction layer for trigger-item, making it degrade into SIM[6]. To reduce the size of parameters, we force TFN share the same embedding layer with Trigger-Aware Net, since embedding layer accounts the most large portion of parameters in CTR models.

Table 1: CTR prediction comparison on the testing set

Method	AUC
Wide & Deep	0.767
DIN	0.791
SIM (TFN)	0.800
DIHN	0.805
TAN	0.806
DIAN	**0.812**
DIAN (remove intention loss)	0.808
DIAN (remove Intention Net)	0.806

Table 2: Intention estimation results on the testing set

Method	Accuracy	AUC
Intention Net	91.3%	0.943

Following the same way as in Eq.4, we generate user interest representation H'_{tar} to the target item. We concatenate three embeddings together and pass them through a MLP layer:

$$\hat{y}_{tfn} = \text{Sigmoid}\left(\text{MLP}(E_{user} \oplus E_{tar} \oplus H'_{tar})\right) \quad (8)$$

where $\hat{y}_{tfn}$ represents the CTR to target item, given user's entering intention is to mini-app, *i.e.*, $P(Click \mid Target, Intent = 0)$.

2.5 Multi-Task Training

Given estimated user intention $\hat{y}_{int}$, and CTRs given user's different intention $\hat{y}_{tan}$ and $\hat{y}_{tfn}$, we can estimate the final CTR as illustrated in Section 2.1 (Eq. 1):

$$\hat{y} = \hat{y}_{int} * \hat{y}_{tan} + (1 - \hat{y}_{int}) * \hat{y}_{tfn} \quad (9)$$

The final objective function consists of CTR loss and intention loss:

$$L = -\frac{1}{|\mathcal{D}|} \sum_{(\hat{y},y) \in \mathcal{D}} \left(y \log \hat{y} + (1-y) \log(1-\hat{y}) \right.$$
$$\left. + \alpha \left(y_{int} \log \hat{y}_{int} + (1 - y_{int}) \log(1 - \hat{y}_{int}) \right) \right) \quad (10)$$

where $\mathcal{D}$ is the training set, $y \in \{0, 1\}$ is the label for CTR prediction task, and $y_{int} \in \{0, 1\}$ is the label for user intention estimation task, *i.e.*, it denotes whether a user's entering intention is the mini-app or trigger item. α is hyper-parameter to balance CTR loss and intention loss, which is set to 0.1 in our experiments.

3 EXPERIMENTS

3.1 Experimental Setup

Dataset: we collect a dataset from Juhuasuan in Taobao, which is a famous mini-app that sells brand-discounted products. We first sample a training set from user impression and click logs from Aug. 1, to Sep. 1, 2022, and then sample the testing set in Sep. 2, 2022, making it strictly simulates online serving. The training set has 10,430,212,219 samples and testing set has 352,959,879 samples, with in total 173,406,246 users and 6,500,833 items involved.

Baselines: We compare DIAN against two types of baselines that are widely used in the industry: 1) Trigger-free methods: Wide & Deep[2], DIN[13], SIM[6]; 2) Trigger-based methods: DIHN[7] and our Trigger-Aware Net (TAN).

Table 3: Online A/B testing results

Method	IPV	PV	CTR
SIM	0.555	23.93	2.32%
DIAN	0.607	24.91	2.43%
Lift rate	+9.39%	+4.10%	+4.74%

Implementation: For all the competitors, dimensions of hidden layers are set to $1024 \times 512 \times 256$. Among sequential models, short-term click sequence is used for DIN and both short- and long-term sequences are used for SIM(TFN), DIHN, TAN and DIAN. Adam optimizer is adopted with learning rate of 0.01. For all attention-based modules, the number of attention heads is set to 4.

3.2 Offline Evaluation

Performance Comparison: Table 1 shows the comparison results of CTR prediction task on the testing set, where we adopt Area Under Curve (AUC) that widely used for industrial recommendation as the measurement metric. The first observation is that trigger-based methods outperforms trigger-free methods, indicating that trigger item that reveals user instant interest is important for TIRA task; Second, among trigger-based methods, DIHN works not very well because it dedicate to improve the relevance between recommendation results and the trigger item, which is undesired for those routine users of mini-app. The last observation is that the proposed DIAN network outperforms all the competitors. We contribute its superior performance to it explicitly identifies user's entering intention thus strike a perfect balance between trigger-free and trigger-based recommendations, making it suitable for TIRA.

Ablation Study: We verify effectiveness of each element of DIAN in Table 1. The first observation is that TFN and TAN can harmoniously work together and their performance can be improved by the Intention Net. This is consistent with the second observation that when we remove the Intention Net and averages outputs of TAN and TFN, AUC value decreases to 0.806. Finally, if we remove intention loss to make DIAN itself balance the outputs of TAN and TFN, the performance also decreases a lot, indicating that it is hard to identify user entering intention within just impression-click samples. The proposed labeling strategy and auxiliary intention estimation task are essential to truly identify user's intention.

User Intention Estimation: Table 2 summarizes the results of intention estimation task, where AUC value reaches 0.943 and accuracy reaches 91.3%, indicating that the proposed Intention Net is effective to identify user's entering intention.

3.3 Online A/B Testing

We implemented and deployed the proposed DIAN for Juhusuan mini-app in Taobao App. From Sep. 16 to Oct. 15, 2022, a strict online A/B test is conducted to compare DIAN against the previous online model SIM. Table 3 shows the comparison results *w.r.t.* Page View (PV), Item Page View (IPV) and CTR, where PV and IPV are the average numbers of items that users browse and click after they enter the mini-app. It is obvious that DIAN outperforms SIM with a large margin: DIAN contributes up to 4.74% lift for CTR, 4.10% lift for PV and 9.39% lift for IPV, which indicates that users not only browse more but also more likely to click items recommended by DIAN.

4 CONCLUSION

In this paper, we define a new task dubbed Trigger-Induced Recommendation in mini-App (TIRA) and identify the key to TIRA is to accurately estimate user's intention when they enter the mini-app. A novel Deep Intention-Aware Network is proposed to achieve this goal. Experiments on a large industrial dataset and online A/B testing verify its effectiveness for the real-world application. Currently, DIAN has been deployed on Taobao App to serve for the full volume traffic of Juhusuan.

REFERENCES

[1] Y. Cao, S. Hu, Y. Gong, Z. Li, Y. Yang, Q. Liu, and S. Ji. Gift: Graph-guided feature transfer for cold-start video click-through rate prediction. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 2964–2973, 2022.

[2] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, et al. Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems*, 2016.

[3] Y. Feng, F. Lv, W. Shen, M. Wang, F. Sun, Y. Zhu, and K. Yang. Deep session interest network for click-through rate prediction. In S. Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 2301–2307. ijcai.org, 2019.

[4] X. Li, C. Wang, J. Tan, X. Zeng, D. Ou, and B. Zheng. Adversarial multimodal representation learning for click-through rate prediction. In Y. Huang, I. King, T. Liu, and M. van Steen, editors, *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, pages 827–836. ACM / IW3C2, 2020.

[5] Z. Lin, H. Wang, J. Mao, W. X. Zhao, C. Wang, P. Jiang, and J. Wen. Feature-aware diversified re-ranking with disentangled representations for relevant recommendation. In A. Zhang and H. Rangwala, editors, *KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14 - 18, 2022*, pages 3327–3335. ACM, 2022.

[6] Q. Pi, G. Zhou, Y. Zhang, Z. Wang, L. Ren, Y. Fan, X. Zhu, and K. Gai. Search-based user interest modeling with lifelong sequential behavior data for click-through rate prediction. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 2685–2692, 2020.

[7] Q. Shen, H. Wen, W. Tao, J. Zhang, F. Lv, Z. Chen, and Z. Li. Deep interest highlight network for click-through rate prediction in trigger-induced recommendation. In F. Laforest, R. Troncy, E. Simperl, D. Agarwal, A. Gionis, I. Herman, and L. Médini, editors, *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*, pages 422–430. ACM, 2022.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, et al. Attention is all you need. *arXiv:1706.03762*, 2017.

[9] R. Xie, R. Wang, S. Zhang, Z. Yang, F. Xia, and L. Lin. Real-time relevant recommendation suggestion. In L. Lewin-Eytan, D. Carmel, E. Yom-Tov, E. Agichtein, and E. Gabrilovich, editors, *WSDM '21, The Fourteenth ACM International Conference on Web Search and Data Mining, Virtual Event, Israel, March 8-12, 2021*, pages 112–120. ACM, 2021.

[10] R. Xie, R. Wang, S. Zhang, Z. Yang, F. Xia, and L. Lin. Real-time relevant recommendation suggestion. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, WSDM '21, page 112–120, New York, NY, USA, 2021. Association for Computing Machinery.

[11] S. Yao, J. Tan, X. Chen, K. Yang, R. Xiao, H. Deng, and X. Wan. Learning a product relevance model from click-through data in e-commerce. In J. Leskovec, M. Grobelnik, M. Najork, J. Tang, and L. Zia, editors, *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, pages 2890–2899. ACM / IW3C2, 2021.

[12] G. Zhou, N. Mou, Y. Fan, Q. Pi, et al. Deep interest evolution network for click-through rate prediction. In *AAAI*, 2019.

[13] G. Zhou, X. Zhu, C. Song, Y. Fan, et al. Deep interest network for click-through rate prediction. In *SIGKDD*, 2018.