Multi-Interest Refinement by Collaborative Attributes Modeling for Click-Through Rate Prediction

Huachi Zhou*
The Hong Kong Polytechnic University
Hung Hom, Hong Kong SAR
huachi.zhou@connect.polyu.hk

Jiaqi Fan
TCL Corporate Research (Hong Kong) Co., Limited
Sha Tin, Hong Kong SAR
garyfan@tcl.com

Xiao Huang
The Hong Kong Polytechnic University
Hung Hom, Hong Kong SAR
xiaohuang@comp.polyu.edu.hk

Ka Ho Li
TCL Corporate Research (Hong Kong) Co., Limited
Sha Tin, Hong Kong SAR
karlli@tcl.com

Zhenyu Tang
TCL Corporate Research (Hong Kong) Co., Limited
Sha Tin, Hong Kong SAR
alantang@tcl.com

Dahai Yu
TCL Corporate Research (Hong Kong) Co., Limited
Sha Tin, Hong Kong SAR
dahai.yu@tcl.com

ABSTRACT

Learning interest representation plays a core role in click-through rate prediction task. Existing Transformer-based approaches learn multi-interests from a sequence of interacted items with rich attributes. The attention weights explain how relevant an item's specific attribute sequence is to the user's interest. However, it implicitly assumes the independence of attributes regarding the same item, which may not always hold in practice. Empirically, the user places varied emphasis on different attributes to consider whether interacting with one item, which is unobserved. Independently modeling each attribute may allow attention to assign probability mass to some unimportant attributes. Collaborative attributes of varied emphasis can be incorporated to help the model more reasonably approximate attributes' relevance to others and generate refined interest representations.

To this end, we novelly propose to integrate a dynamic collaborative attribute routing module into Transformer. The module assigns collaborative scores to each attribute of clicked items and induces the extended Transformer to prioritize the influential attributes. To learn collaborative scores without labels, we design a diversity loss to facilitate score differentiation. The comparison with baselines on two real-world benchmark datasets and one industrial dataset validates the effectiveness of the framework.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; *Personalization*; *Learning to rank*.

*This research is conducted during an internship at TCL Corporate Research (Hong Kong) Co., Limited.

CIKM '22, October 17–21, 2022, Atlanta, GA, USA

ACM ISBN 978-1-4503-9236-5/22/10…$15.00
https://doi.org/10.1145/3511808.3557652

KEYWORDS

Click-through Rate Prediction; Multi-Interest; Attention-Smoothing

ACM Reference Format:
Huachi Zhou, Jiaqi Fan, Xiao Huang, Ka Ho Li, Zhenyu Tang, and Dahai Yu. 2022. Multi-Interest Refinement by Collaborative Attributes Modeling for Click-Through Rate Prediction. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22), October 17–21, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3511808.3557652

1 INTRODUCTION

Learning interest representations from concrete behaviors has received broad attention in the click-through rate prediction task. Its goal is to predict items to be interacted next from the candidate item set and rank them in line with user preferences. Conventional works statically concatenate item representations and input them into multi-layer perceptron to predict target item click probability [3, 8]. They often leave extracting high-level semantics from behavior sequences to the embedding layers without explicit modeling. Thus the concatenated vector can not accurately describe user personalized interests. Many recent efforts leverage candidate items to guide model to actions of interests and summarize them in a fixed-length vector [5, 14].

Users' interests are diverse. A user may interact with many conceptually different products at a time [9]. Representing user interests with a single vector irreversibly bottlenecks the expressive information from the diversified user action sequences. Thus, this line of works often aims to learn multiple latent interests of each user. An item is associated with several attributes, and rich attributes bring opportunities to multi-interest mining. The Transformer-based [11] and capsule-based framework [2] are successively designed to capture multiple diverse interests expressed by the user behaviors. The attention weight explains the relevance between attributes in each attribute sequence. Each attribute is represented by the attentive sum of others in the sequence based on the attention weights, and then each processed sequence is pooled into one interest respectively.

Despite the success of the attention mechanism in multi-interest click-through rate prediction, it implicitly assumes the independence of attributes regarding the same item and may be insufficient

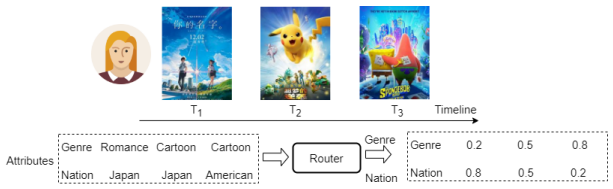

Figure 1: Each played movie is associated with two attributes: movie genre and nation. The router models collaborative attributes and outputs scores.

to achieve further satisfactory model capacity. Empirically, the user places varied emphasis on different attributes to consider whether interacting with one item, which is unobserved. As illustrated in Figure 1, the user interacted with three movies, each of which associated with two attributes, movie genre and nation. Traditional efforts may assume no correlation between movie genre and nation at each time step and adopt two attention heads to process two attribute sequences. Nevertheless, we could observe that user preferences are skewed toward cartoons and Japanese movies. The router models collaborative scores and considers each attribute contributions driving users to interact with an item. The attention could leverage collaborative scores to more reasonably approximate attributes' relevance to others and generate refined interest representations.

Motivated by aforementioned problems, we explore modeling collaborative attributes vertically for each click. The collaborative scores represent the confidence of each attribute being the reason why the user consumes this item. It pushes attention to focus on influential attributes, which is beneficial to refine the quality of the interest representations. However, two challenges ensue. First, each click is often induced by intent synthesis, but the underlying intent contributions are unobservable, meaning there is no available collaborative scores as a supervision signal to assist score assignment. Second, the assigned collaborative scores for each attribute at the same time step may collapse to a uniform distribution, hindering the efforts to highlight the heterogeneous contributions of attributes.

To this end, we propose a novel framework CAMN (Collaborative Attributes Based Multi-Interest Network) for click-through rate prediction. We summarize our contributions as follows:

- To better approximate attributes' relevance to others and generate refined interests, we explore explicitly quantifying the importance of attributes driving user to interact with an item and propose a novel framework CAMN.
- We propose a dynamic collaborative attribute routing module to estimate collaborative scores and a diversity loss to drive the model to learn imbalanced score distributions.
- Extensive experiments on public and industrial datasets prove CAMN effectiveness.

2 COLLABORATIVE ATTRIBUTES BASED MULTI-INTEREST NETWORK

Notations: We introduce model framework notations in detail. Let $\mathcal{I}$ represent the item set collected from all sessions. Each item $i \in \mathcal{I}$ is associated with k attributes (including item identifier), which are a small subset of attribute set $\mathcal{A}$. Given a sequence s with $[a_1^s, a_2^s \dots a_l^s]$ with l items, the click-through rate prediction task is to estimate the probability of the next item being clicked. The r-th corresponding attribute sequence associated with each item in s could be represented as $[a_{r,1}^s, a_{r,2}^s \dots a_{r,l}^s]$[1]. To project k consecutive attributes for one item from sparse vector to dense vector, we look up attribute embedding matrix $\boldsymbol{E} \in \mathbb{R}^{|\mathcal{A}| \times d}$ and concatenate them together, where $|\cdot|$ is the cardinality and d is embedding dimension. Following this way, we obtain the composite item embedding matrix $\boldsymbol{X} \in \mathbb{R}^{l \times kd}$ and elementary r-th attribute embedding matrix $\boldsymbol{X}_r \in \mathbb{R}^{l \times d}$ for sequence s.

2.1 Collaborative Attribute Routing Module

To collaboratively model attribute importance, we analyze user intent and output collaborative scores for all attributes. A simple yet effective collaborative attribute routing module is introduced to estimate score distribution, which is further utilized to adjust attribute relevance in Transformer. Given input matrix $\boldsymbol{X}$, layer normalization [1] first rescales all the hidden units and prevents embedding norm from influencing the stability of model training:

$$\tilde{\boldsymbol{X}} = Layernorm(\boldsymbol{X}). \tag{1}$$

A trainable matrix $\boldsymbol{W}_h \in \mathbb{R}^{kd \times k}$ is utilized to reduce the dimension to the number of attributes k and produce the initial routing scores $\tilde{\boldsymbol{P}}$. A user may only consider a few attributes to facilitate this interaction. So we only select top q elements and set the rest to $-\infty$. Then we normalize collaborative attribute scores independently at each time step to constrain them in [0,1] and sum to 1. The final collaborative score matrix $\boldsymbol{P} \in \mathbb{R}^{l \times k}$ is formulated as:

$$\boldsymbol{P} = Softmax(keepTopq(\tilde{\boldsymbol{X}}\boldsymbol{W}_h, q)), \tag{2}$$

and

$$keepTopq(\tilde{\boldsymbol{P}}_i, q)_j = \begin{cases} \tilde{\boldsymbol{P}}_{i,j} & \text{if } \tilde{\boldsymbol{P}}_{i,j} \text{ in the top q element} \\ -\infty & \text{otherwise} \end{cases}. \tag{3}$$

Notably, only the top q values have nonzero derivatives with respect to the weights of the module.

2.2 Modified Transformer

We leverage Transformer [10] to produce relevance scores between instances in the r-th attribute sequence. Scores obtained in collaborative attribute routing module are incorporated to adjust attribute relevance and then help refine downstream interest representations. Attributes allocated high scores are more likely the reason why the user would like to interact with the item. Correpondingly, the r-th attribute sequence should receive more attention from the model. The original weights are obtained by the multiplication between latent item representations, and we apply collaborative score matrix $\boldsymbol{P}$ to scale the magnitude of the obtained weights. This process outputs smoothed attention weight matrix $\boldsymbol{A}_r$ defined by:

$$\boldsymbol{A}_r = Softmax(\boldsymbol{P} \odot (\frac{\boldsymbol{X}_r\boldsymbol{W}_r^Q(\boldsymbol{X}_r\boldsymbol{W}_r^K)^\top}{\sqrt{d}})), \tag{4}$$

[1]We omit superscript s in the following notations for brevity. And all item attributes should be in the same order.

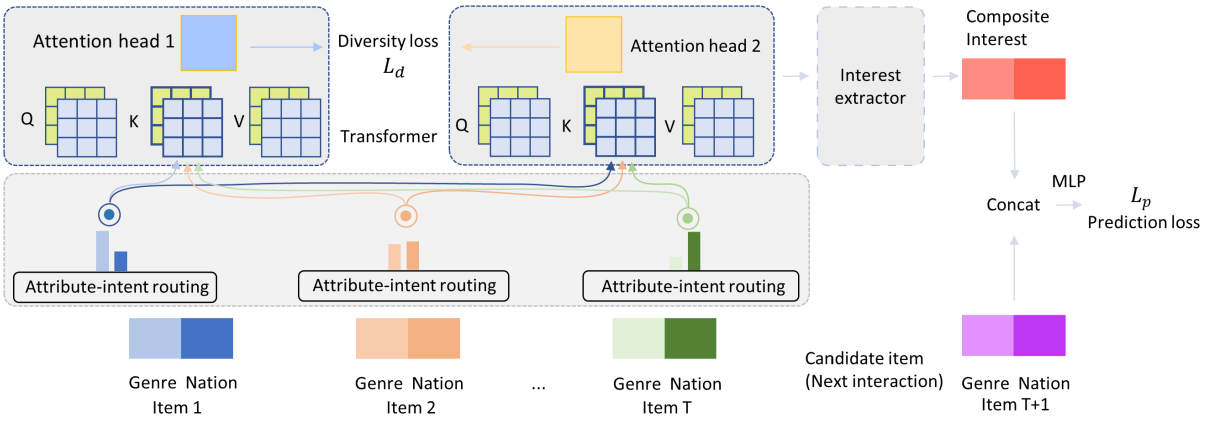

Figure 2: Toy example processed by CAMN pipeline. Shades of color represent different attributes.

where learnable parameter query, key matrices $\boldsymbol{W}_r^Q, \boldsymbol{W}_r^K \in \mathbb{R}^{d\times d}$ are separately to map input to different latent spaces under r-th attention head.

Along this line, attention weights among attributes in the r-th attribute sequence are re-evaluated. The attribute assigned higher scores will have more relevance with others and take a more dominant position in the joint representation:

$$\hat{\boldsymbol{X}}_r = \boldsymbol{A}_r(\boldsymbol{X}_r\boldsymbol{W}_r^V), \tag{5}$$

where $\boldsymbol{W}_r^V \in \mathbb{R}^{d\times d}$ represents parameter value matrix. To integrate non-linear signal into item representations, we apply feed-forward neural network to item representations:

$$\tilde{\boldsymbol{X}}_{\boldsymbol{r}} = max(0, \hat{\boldsymbol{X}}_r\boldsymbol{W}_1 + \boldsymbol{b}_1)\boldsymbol{W}_2 + \boldsymbol{b}_2, \tag{6}$$

where $\boldsymbol{W}_1, \boldsymbol{W}_2 \in \mathbb{R}^{d\times d}$, $\boldsymbol{b}_1, \boldsymbol{b}_2 \in \mathbb{R}^d$ are trainable matrices and biased terms, respectively.

Following previous practices in DMIN [11], $\tilde{\boldsymbol{X}}_{\boldsymbol{r}}$ is pooled by interest extractor to learn refined interest embedding. Finally, k interest vectors and target item embedding are concatenated together and sent to MLP(Multi-layer Perceptron) to estimate the clicked probability $\tilde{y}$.

2.3 Diversity Loss

Only multiplying collaborative scores may not sufficiently push the model to differentiate attribute contributions at each click. To distinguish vital attribute driving user behaviors, we design a diversity learning objective to promote imbalanced collaborative score assignment among all attributes. The loss function serves as a regularization technique and encourages the diversity of information attended by different attention heads. It is given by the sum of element-wise multiplication between each pair of attention weight matrix of k attributes:

$$\mathcal{L}_d = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\left\|\boldsymbol{A}_i \odot \boldsymbol{A}_j\right\|. \tag{7}$$

The model is jointly optimized by the diversity loss, and the next-item clicked probability prediction loss. Our final training target is to minimize the following objective:

$$\mathcal{L} = \alpha\mathcal{L}_d + \mathcal{L}_p, \tag{8}$$

Table 1: Dataset Statistics.

Dataset	Users	Items	Attributes	Samples
Books	1686577	1861983	6	3373154
Industry	20448	7476	5	1525785
Electronics	651680	337652	6	1303360

where α controls the weight of diversity loss. We treat the recommendation task as a classification problem. The next-item clicked probability prediction loss concerning the target item with truth label y is given by:

$$\mathcal{L}_p = -[y\log\tilde{y} + (1-y)\log(1-\tilde{y})]. \tag{9}$$

3 EXPERIMENTS

In this section, we conduct experiments over two benchmark datasets and one industrial dataset to answer three questions.

- How effective is CAMN compared with the state-of-the-art baselines?
- How much does each component of CAMN contribute?

3.1 Dataset Preprocessing

We evaluate all models on two publicly available datasets and one industrial dataset. We summarize the dataset statistics in Table 1. Notably, Amazon Books and Amazon Electronics released by McAuley et al. [7] contain product reviews and metadata from the Amazon website. The Industrial dataset contains movie play records collected from a leading home TV recommendation platform. For each user, we obtain their reviews and sort them by ascending timestamps. The last interaction is treated as positive example, and we randomly sample an item from the whole item set and pair it with the user as a negative example. We randomly select 80% users in each dataset as training set, 10% as validation set, and 10% as test set. The maximum length of sequences is set as 20.

3.2 Experimental Setup

To enable a fair comparison, we use the source code released by authors for the baseline. We implement CAMN with Tensorflow and learning rate is set as 0.002. Following previous experimental

protocols [11, 12], the number of embedding dimension d equals 18 and Adam optimizer is utilized to optimize all models. Batch size equals 2048 and α is 0.0001. The number of interests is aligned with the attribute number. AUC metric (Area Under Curve) [4] is incorporated to evaluate all models. We repeatedly run all models five times and report average results and standard deviation.

3.3 Compared Methods

We incorporate three groups of representative baseline methods to verify our model performance. First, to verify the effectiveness of capturing dynamic user interest, we include two models, **Wide&Deep** [3] and **PNN** [8] which ignore chronological order of clicked items. Second, to prove the superiority of multiple user interests, we include succeeding baselines **DIEN** [13] and **DHAN** [12], which focus on learning single interest representation. Third, to demonstrate the usefulness of modeling collaborative attributes in interest refinement, we include the following methods **ComiRec-SA** [2], **MIND** [6], and **DMIN** [11], which aim to extract multiple interests from user sequences.

3.4 Comparison with Baselines

We start to answer the first problem of how effective CAMN is compared with the state-of-the-art baselines. From Table 2, we have several observations.

Wide&Deep and PNN do not achieve satisfactory results compared with other baselines. They statically concatenate item embedding together and do not model item transition in the sequence. Thus user interest is not explicitly revealed and they fail to exactly predict the possible next item. Dynamic user interest modeling methods, including DIEN and DHAN exhibit competitive performance with multi-interest models. They achieve better results than the first group baseline. The improvement shows that learning interest representation is beneficial for click-through rate prediction task.

Different multi-interest models have varied abilities in dealing with user complex interests. The first and second group baselines lose to DMIN. It demonstrates the advantage of extracting multi-interest than learning a unified interest. The third group baselines are also inferior to DMIN, which indicates the superiority of Transformer in multi-interest modeling.

CAMN achieves consistent improvement over all the baselines on three datasets which validates the effectiveness of our framework. Remarkably, the improvement over DMIN proves that the quality of interest representation in CAMN is better than DMIN. Because collaborative attributes help analyze user intent and adjust attention weight scores in each attribute sequence. It indicates the effectiveness of collaborative attributes refining multi-interest.

3.5 Ablation Study

We turn to investigate the second problem of how much each component of CAMN contributes to the whole model performance. Now we introduce two variants of CAMN: CAMN-route and CAMN-loss. CAMN-route is obtained by CAMN excluding diversity loss which is used to validate the effectiveness of learning imbalanced collaborative score distribution in each click. CAMN-loss is obtained by CAMN removing collaborative attribute routing module, which is derived to validate the effectiveness of modeling collaborative attributes. The ablation study results on three datasets are summarized in Table 3.

Table 2: Comparison with baselines w.r.t. AUC scores.

Methods	Books	Industry	Electronics
	AUC	AUC	AUC
Wide&deep	0.6594 ± 0.0025	0.7525 ± 0.0007	0.7057 ± 0.0020
PNN	0.6695 ± 0.0036	0.7565 ± 0.0003	0.7216 ± 0.0005
DIEN	0.7098 ± 0.0005	0.7596 ± 0.0001	0.7352 ± 0.0004
DHAN	0.7076 ± 0.0029	0.7637 ± 0.0000	0.7339 ± 0.0001
MIND	0.6710 ± 0.0028	0.7574 ± 0.0002	0.7232 ± 0.0006
Comirec-SA	0.6681 ± 0.0087	0.7620 ± 0.0001	0.7267 ± 0.0003
DMIN	<u>0.7104</u> ± 0.0015	<u>0.7659</u> ± 0.0001	<u>0.7344</u> ± 0.0007
CAMN	**0.7138** ± 0.0004	**0.7698** ± 0.0001	**0.7376** ± 0.0001

Table 3: Ablation study.

Methods	Books	Industry	Electronics
	AUC	AUC	AUC
CAMN-loss	0.7121 ± 0.0010	0.7659 ± 0.0001	0.7325 ± 0.0005
CAMN-route	0.6931 ± 0.0032	0.7689 ± 0.0001	0.7341 ± 0.0009
CAMN	**0.7138** ± 0.0004	**0.7698** ± 0.0001	**0.7376** ± 0.0001

From the results, we can observe that CAMN outperforms its two variants. We infer that without diversity loss, scores generated by collaborative attribute routing module are more evenly distributed and model struggles to distinguish decisive attribute from all associated ones. Without collaborative attribute routing module, diversity loss may destroy attention weight matrix structure and negatively affect model performance. The gap between CAMN and its variants validates the effectiveness of the proposed collaborative attribute routing module and diversity loss.

3.6 Conclusion

In this paper, we propose a novel model named CAMN to refine multi-interest representation by modeling collaborative attributes for click-through rate prediction. Collaborative attribute routing module and diversity loss are introduced. The collaborative attribute routing module outputs intent scores among all attributes to rescale attention weight, thus leading to refined interest representation. The diversity loss promotes differentiating collaborative scores on each attribute concerning one item. Empirical results across three real-world datasets validate the effectiveness of CAMN. Our future works are to explore the possibility of diversity recommendation based on learned collaborative scores.

ACKNOWLEDGMENTS

The authors gratefully acknowledge receipt of the following financial support for the research, authorship, and/or publication of this article. This work was supported in part by the Hong Kong Polytechnic University, Start-up Fund (project number: P0033934).

REFERENCES

[1] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. 2016. Layer normalization. *arXiv preprint arXiv:1607.06450* (2016).

[2] Yukuo Cen, Jianwei Zhang, Xu Zou, Chang Zhou, Hongxia Yang, and Jie Tang. 2020. Controllable multi-interest framework for recommendation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2942–2951.

[3] 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.

[4] Tom Fawcett. 2006. An introduction to ROC analysis. *Pattern recognition letters* 27, 8 (2006), 861–874.

[5] 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).

[6] Chao Li, Zhiyuan Liu, Mengmeng Wu, Yuchi Xu, Huan Zhao, Pipei Huang, Guoliang Kang, Qiwei Chen, Wei Li, and Dik Lun Lee. 2019. Multi-interest network with dynamic routing for recommendation at Tmall. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 2615–2623.

[7] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. 2015. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*. 43–52.

[8] Yanru Qu, Han Cai, Kan Ren, Weinan Zhang, Yong Yu, Ying Wen, and Jun Wang. 2016. Product-based neural networks for user response prediction. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*. IEEE, 1149–1154.

[9] Qiaoyu Tan, Jianwei Zhang, Jiangchao Yao, Ninghao Liu, Jingren Zhou, Hongxia Yang, and Xia Hu. 2021. Sparse-interest network for sequential recommendation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 598–606.

[10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[11] 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.

[12] 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.

[13] 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.

[14] 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.