

# Graph Intention Network for Click-through Rate Prediction in Sponsored Search

Feng Li, Zhenrui Chen, Pengjie Wang<sup>†</sup>, Yi Ren, Di Zhang, Xiaoyu Zhu

Alibaba Group

{adam.lf,zhenrui.cZR,pengjie.wpj,hengrui.ry,di.zhangd,benjamin.zxy}@alibaba-inc.com

## ABSTRACT

Estimating click-through rate (CTR) accurately has an essential impact on improving user experience and revenue in sponsored search. For CTR prediction model, it is necessary to make out user's real-time search intention. Most of the current work is to mine their intentions based on users' real-time behaviors. However, it is difficult to capture the intention when user behaviors are sparse, causing the ***behavior sparsity*** problem. Moreover, it is difficult for user to jump out of their specific historical behaviors for possible interest exploration, namely ***weak generalization*** problem. We propose a new approach Graph Intention Network (GIN) based on co-occurrence commodity graph to mine user intention. By adopting multi-layered graph diffusion, GIN enriches user behaviors to solve the behavior sparsity problem. By introducing co-occurrence relationship of commodities to explore the potential preferences, the weak generalization problem is also alleviated. To the best of our knowledge, the GIN method is the first to introduce graph learning for user intention mining in CTR prediction and propose end-to-end joint training of graph learning and CTR prediction tasks in sponsored search. At present, GIN has achieved excellent offline results on the real-world data of the e-commerce platform outperforming existing deep learning models, and has been running stable tests online and achieved significant CTR improvements.

## CCS CONCEPTS

- **Information systems** → **Sponsored search advertising; Recommender systems.**

## KEYWORDS

sponsored search, click-through rate prediction, graph neural network, intention mining

### ACM Reference Format:

Feng Li, Zhenrui Chen, Pengjie Wang, Yi Ren, Di Zhang, Xiaoyu Zhu. 2019. Graph Intention Network for Click-through Rate Prediction in Sponsored Search. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19), July 21–25, 2019, Paris, France*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3331184.3331283>

<sup>†</sup>This author is the one who gives a lot of guidance in the work.

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](mailto:permissions@acm.org).

SIGIR '19, July 21–25, 2019, Paris, France

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6172-9/19/07...\$15.00

<https://doi.org/10.1145/3331184.3331283>

## 1 INTRODUCTION

In sponsored search, estimating click-through rate accurately is essential to improve revenue and user experience. For accurate estimation of CTR, It is critical to understand user's real-time search intentions in the CTR prediction task, because the majority of users do not describe their search intention completely through query.

Currently, lots of user intention mining method is proposed. Temporal Deep Structured Semantic Model (TDSSM) [4] characterizes user's intention as long-term and short-term to capture their preference and real-time intention. Dynamic REcurrent bAsket Model (DREAM) [8] uses recurrent neural network (RNN) to model user's behavior sequence to improve user intention expression. Furthermore, Deep Interest Network (DIN) [9] indicates that user interest is diverse, and uses the attention mechanism to calculate the relevance between the current advertising commodity and historical commodities clicked by the user.

However, these intention recognition methods mentioned above mainly focus on user's historical behaviors, i.e., user's intention is summarized according to historical behaviors. This kind of methods have two disadvantages: ***behavior sparsity*** and ***weak generalization***. Behavior sparsity means that it is difficult to capture the user's real-time intention when user's behavior is sparse. Weak generalization refers to the user's inability to jump out of their specific historical behavior for possible interest exploration.

In addition, some graph embedding methods are introduced into the CTR prediction task by a two-stage approach. [6] uses DeepWalk [2] to generate node sequence and the Skip-Gram model is used for graph embedding. Then, the learned node representation is further used in the CTR predict task. There are numerous work proposed for graph embedding. Graph Convolutional Network (GCN) [1] aggregates neighbor nodes through mean-pooling and generates new representations with the current nodes through nonlinear functions. Graph Attention Network (GAT) [5] further proposed attention-based neighbor aggregation by calculating the correlation between the current node and neighbors.

These graph embedding based methods have achieved significant results, but these methods are not directly optimized for specific CTR prediction task, which means that these methods firstly learn graph node representation by unsupervised or semi-supervised methods and then use the learned node representations to predict the CTR. This kind of training methods is not optimized for the final goals, and node representations are not adjusted by the specific tasks, thus becoming the bottleneck of the expression ability in the CTR prediction task.

We propose a new approach Graph Intention Network (GIN) based on co-occurrence commodity graph to solve these problems. Firstly, the GIN method enriches user's behavior by multi-layered

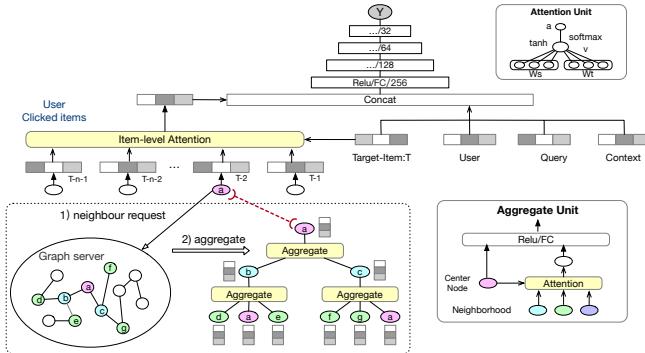
graph diffusion of user historical behaviors, and solves the behavior sparsity problem. Secondly, the weak generalization problem is alleviated by introducing co-occurrence relationship of commodities to explore the potential preferences of users. Finally, we combine this intention mining method based on co-occurrence commodity graph with the CTR prediction task by end-to-end joint training.

The main contributions of this paper are as follows.

- (1) The end-to-end joint training of graph learning and CTR prediction tasks is proposed for the first time in the sponsored search ranking model.
- (2) The behavior sparsity and weak generalization problems are alleviated by the multi-layered intention diffusion and aggregation based on the co-occurrence click relationship graph.
- (3) The effectiveness of the proposed GIN method is verified by offline and online experiments.

## 2 THE PROPOSED APPROACH

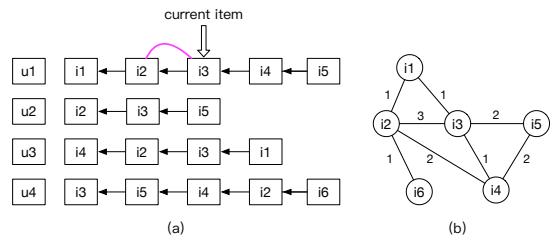
In this section, we introduce the GIN method in detail, as shown in Fig. 1. Firstly, the construction of co-occurrence commodity graph based on historical behaviors is introduced. Secondly, how to diffuse and aggregate multiple layers of implicit intention is introduced based on the co-occurrence commodity graph. Finally, the end-to-end joint training method is presented to combine the graph-based intention mining with CTR prediction tasks.



**Figure 1:** The proposed end-to-end joint training method combines graph-based intention mining with CTR prediction tasks. Each historical clicked sample first performs a multi-layered neighbor query on the graph service, and the attention mechanism is used to perform neighbor aggregation according to correlations between the current node and the neighbor nodes. Finally, the aggregated intention results and other features are concatenated as inputs for CTR prediction.

### 2.1 Graph construction

User historical clicks are regarded as a sequence, only click behaviors in the last month was intercepted to balance performance and effectiveness. The behavior sequences are segmented into sessions based on query similarity to prevent edge construction across dissimilar queries. Each commodity in the session constructs several undirected edges by window size, thus constructing a co-occurrence commodity graph. The node type is commodity only, and the weight of edge indicates the number of co-occurrence times.

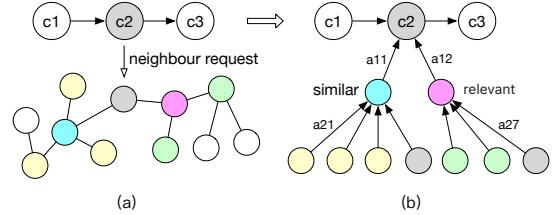


**Figure 2:** The graph is constructed based on user history behaviors. (a) Each row represents a user's click sequence. The black arrow indicates the behavior direction, and the red arrow indicates the graph edge when the window size is 1. (b) In the co-occurrence commodity graph, nodes represent clicked commodities, and edge weights indicate the numbers of co-occurrence clicks.

The detailed graph construction is shown in the Fig. 2. Assuming the window size is 1, we construct an undirected edge to the left of each node in the sequence, and the co-occurrence commodity graph is obtained after processing each user's click sequence.

### 2.2 Intention diffusion and aggregation

We diffuse user's behavior sequence on co-occurrence graph to enrich user's intention expression as shown in Fig. 3. Fig. 3(a) contains user's behavior and co-occurrence commodity graph. Fig. 3(b) is obtained by performing multi-layered neighbor diffusion on graph for each commodity of user click sequence. Then, The attention mechanism is applied to aggregate the tree-like intention.



**Figure 3:** The multi-layered intention diffusion and aggregation process is applied based on the co-occurrence commodity graph. Here  $c_1$ ,  $c_2$  and  $c_3$  represent user's click sequence. (a) indicates that the sequence of behavior is multi-layer diffusion into the graph. (b) indicates that the results of multi-layer diffusion are aggregated using the attention mechanism.

Diffusing user's real-time behavior on co-occurrence graph can recall two kinds of commodity. One is extremely similar commodities in same behavior cluster, it enriches user's behavior which is benefit for solving behavior sparsity problem. The other is relevant but not extremely similar commodities in another behavior cluster, which help user to jump out of their specific historical behavior for possible interest exploration, so the weak generalization problem is alleviated. Similar and relevant commodity is further described in Fig. 4.

The intention diffusion and aggregation process is further detailed in Algorithm 1 inspired by [5, 7]. We first diffuse each commodity of the user's click sequence in layers to explore commodities that have a strong co-occurrence relationship with the user's current click. Then aggregate the diffused commodities layer by



**Figure 4:** The left and right parts represent two different behavior clusters with similar commodities in the co-occurrence commodity graph. The connection between cluster 1 and cluster 2 indicates strong relationship between two clusters, and the system utilizes these connections to help user jump out from cluster 1 to cluster 2 which is a potential preference.

layer from the outermost layer with AGGREGATE function described in Algorithm 2. Finally, we select the commodities of the user’s click sequence by attention mechanism to increase the weight of the relevant commodities, and finally obtains the diverse vector including user’s potential preference.

#### Algorithm 1 Graph Intention Discovery(GID)

**Input:** Current ranking ad  $ad$ , User click behaviors  $pre\_clicks$ , Depth  $K$ ; Commodity similarity graph  $G$ ; Neighbor select function  $N$ ;

**Output:** User implicit intention embedding  $uii$

```

1: /* Implicit intention propagation */
2:  $S^{(K)} \leftarrow pre\_clicks$ 
3: for  $k = K \rightarrow 1$  do
4:    $S^{(k-1)} \leftarrow S^{(k)}$ 
5:   for  $u \in S^{(k)}$  do
6:      $S^{(k-1)} \leftarrow S^{(k-1)} \cup N_G(u)$ 
7:   end for
8: end for
9: /* Implicit intention aggregation */
10:  $h_u^{(0)} \leftarrow x_u, \forall u \in S^{(0)}$ 
11: for  $k = 1 \rightarrow K$  do
12:   for  $v \in S^{(k)}$  do
13:      $H \leftarrow \{h_u^{(k-1)}, u \in N_G(v)\}$ 
14:      $h_v^{(k)} \leftarrow \text{AGGREGATE}^{(k)}(h_v^{(k-1)}, H)$ 
15:   end for
16: end for
17: /* Generate user intention embedding by attention */
18:  $a_c \leftarrow \text{softmax(score}(h_{ad}, h_c^{(K)})\text{), } \forall c \in S^K$ 
19:  $uii \leftarrow \sum_{c \in S^{(K)}} a_c h_c^{(K)}$ 

```

#### Algorithm 2 AGGREGATE

**Input:** Current node embedding  $h_u$  for node  $u$ , Set of neighbor embeddings  $H = \{h_v, v \in N_u\}$ , Symmetric vector function  $\gamma(\cdot)$

**Output:** AGGREGATED result  $h_u^{agg}$  for node  $u$

```

1:  $\alpha_{uv} = \frac{\exp(\text{ReLU}(z^T \cdot [W h_u || W h_v]))}{\sum_{k \in N_u} \exp(\text{ReLU}(z^T \cdot [W h_u || W h_k]))}, \forall v \in N_u$ 
2:  $n_u \leftarrow \gamma(\{\text{ReLU}(M h_v + m) | v \in N_u\}, \alpha_u)$ 
3:  $h_u^{agg} \leftarrow \text{ReLU}(B \cdot \text{CONCAT}(h_u, n_u) + b)$ 

```

### 2.3 End-to-end joint training method

The end-to-end joint training framework with graph-based intention mining and CTR prediction is shown in Fig. 1. Firstly, we construct a co-occurrence commodity graph based on the method of

Section 2.1, and use the graph engine euler[3] to build a real-time graph neighbor query service. During the training phase, the multi-layered neighbor query is performed on the graph for each item in user’s click sequence, and then the neighbors are aggregated according to the method described in 2.2 to obtain the intention vector. Secondly, this vector is concatenated with other features (e.g. query, user, ad and its statistical ctr) for CTR prediction.

In this process, the neighbor query, aggregation in graph and forward propagation are carried out in an end-to-end manner. Representation of graph node is updated by the back propagation algorithm based on the cross entropy loss defined in equation 1. The forward propagation process is further detailed in Algorithm 3.

---

#### Algorithm 3 Graph Intention Network

---

**Input:** Set of samples with  $(query, user, ad, pre\_clicks)$ ; depth Parameter  $K$ ; Forward propagation function  $forward$ ; Commodity similarity graph  $G$ ; Neighbor select function  $N$ ;

**Output:** Prediction of click-through rate  $pctr$

```

1:  $h \leftarrow \text{GID}(ad, pre\_clicks, K, G, N)$ 
2: features  $\leftarrow \text{CONCAT}(\{h_{query}, h_{user}, h_{ad}, h\})$ 
3:  $pctr \leftarrow \text{sigmoid}(forward(\text{features}))$ 

```

---

**Loss function:** The objective function of the joint training method is the cross entropy loss function as follows:

$$L = -\frac{1}{N} \sum_{i=0}^N y_i \log(pctr_i) + (1 - y_i) \log(1 - pctr_i) \quad (1)$$

Where  $N$  is the total number of samples,  $y_i$  is the label of the  $i$ th sample, and  $pctr_i$  is the GIN forward propagation of the  $i$ th sample.

## 3 EXPERIMENTS

To evaluate the performance of the proposed GIN method in CTR prediction tasks, we designed offline comparison experiments and further verified through online A/B testing.

### 3.1 Experimental Setup

**Graph data:** The co-occurrence commodity graph is constructed using users’ click behavior data during 30 days. There are 1 billion nodes and 8 billion edges. Types of graph node are all commodities. And the average output degree of graph node is 4.

**Train and Test data:** Train data contains about 14 billion samples. Another 2 billion unseen samples are used to assess the performance of different CTR prediction models. Features include sparse id features and statistical features, corresponding to query, user, commodity, and historical behaviors.

**Competitors:** We conduct experiments with several competitive methods on CTR modeling. (1) **Base**: the baseline model for large scale CTR prediction task is neural factorization machines (NFM) which is widely used in industrial product. In this model, the sequence of user behavior is aggregated into an intention vector by sum-pooling. (2) **DIN**: This model uses the attention mechanism to weight the user behavior commodities and obtains the representation of user intention. (3) **GIN**: The proposed method combines graph intention mining with CTR prediction task. The length of previous clicks is 20, and the depth parameter  $K$  is set to 2. A 5-layer full-connection perceptron is adopted as the forward

**Table 1: Comparison of effects of different models.**

Method	Delta AUC
DIN	+0.24%
GIN(ours)	<b>+0.60%</b>

**Table 2: Comparison of different neighbor numbers.**

#Neighbors	Delta AUC	Time Cost
GIN-5	+0.39%	7h
GIN-10	+0.52%	12h
GIN-20	<b>+0.60%</b>	20h

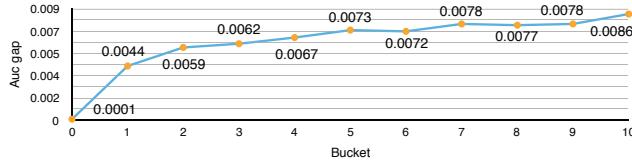
network with ReLU nonlinear activation. The neighbor is selected by the Top- $N$  function according to the edge weight.

### 3.2 Offline evaluation

The AUC is adopted as the offline performance metric. Higher AUC demonstrate better ranking performance. Same train and test data are used in these three CTR prediction models (Base, DIN, GIN). The model effect is obtained after model parameters and optimizer configuration are all optimized. Note that a 0.001 AUC increment means significantly performance improvement in our scenario.

The experimental results are shown in Table 1. Compared with DIN and Base, GIN has a significant effect increment. DIN provides diverse intention expression through the attention mechanism, which improves the model ability to capture user's intention. GIN further introduces the implicit intention information with graph diffusion, and solves the problems of behavior sparsity and weak generalization, which achieves the best CTR prediction performance.

Effect of GIN using different behavior lengths are shown in Fig. 5. The bucket id indicates different behavior lengths. AUC gap indicates GIN outperforms NFM significantly. The 0th bucket indicates GIN cannot perform the effect when there is no historical clicks. For the case with less historical clicks, the effect of GIN has improved slightly, indicating that GIN has an effect on enriching user intention expression. With more historical clicks, the improvement of GIN is more obvious. The reason may be that the user's intention is richly expressed. At the same time, it can discover user's potential preference to help user to migrate.

**Figure 5: Comparison of different behavior lengths.**

**Neighbor number:** In order to further explore the effect of the neighbor number on the model and the impact of each epoch time consumption, we compare the AUC changes and training time cost where the neighbor number is set as 3, 5, 10, 20 under K=2 conditions. As shown in Table 2, as the number of neighbors increases, the effect becomes better, and the time cost also increases linearly. The reason is that the network overhead of the distributed system increases as the number of neighbors increases.

**Neighbor depth:** We compare AUC and time cost for these CTR prediction models to explore the effect of intention diffusion

**Table 3: Comparison of different neighbor depth.**

Hop	Delta AUC	Time Cost
GIN-order 1	+0.45%	<b>8h</b>
GIN-order 2	<b>+0.60%</b>	20h

**Table 4: Comparison of online CTR for 3 consecutive days.**

model	T	T+1	T+2
DIN	+0.65%	+0.66%	+0.50%
GIN	<b>+1.46%</b>	<b>+1.82%</b>	<b>+1.67%</b>

within different neighbor depth as shown in Table 4. GIN-0 means no neighbor info is utilized and GIN-2 means neighbors within two hop are aggregated. The AUC gap increases greatly as the depth grows, while time cost also increases rapidly.

### 3.3 Online A/B Test

We designed an online A/B test to further evaluate the performance of GIN. The comparisons of online CTR for different models during 3 consecutive days are shown in Table 4. The average CTR of GIN increased by 1.65%, indicating that GIN can effectively improve the effect of the CTR prediction task.

## 4 CONCLUSION

In this paper, we propose a novel approach GIN for CTR prediction in sponsored search. Using the end-to-end joint learning method of co-occurrence commodity graph and CTR prediction task, two important problems in user intention mining, i.e., behavior sparsity and weak generalization, are solved through the diffusion and aggregation of historical behaviors. Experiments on offline and online real-world dataset demonstrate the proposed GIN achieved excellent performance.

## REFERENCES

- [1] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in neural information processing systems*. 3844–3852.
- [2] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 701–710.
- [3] Yi Ren, Siran Yang, Yan Zhang, Yuan Wei, Genbao Chen, Xu Tian, Shuai Li, and Di Zhang. 2018. euler. <https://github.com/alibaba/euler/>.
- [4] Yang Song, Ali Mamdouh Elkahky, and Xiaodong He. 2016. Multi-rate deep learning for temporal recommendation. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 909–912.
- [5] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).
- [6] Jizhe Wang, Pipei Huang, Huan Zhao, Zhibo Zhang, Binqiang Zhao, and Dik Lun Lee. 2018. Billion-scale commodity embedding for e-commerce recommendation in alibaba. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 839–848.
- [7] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. 2018. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 974–983.
- [8] Feng Yu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2016. A dynamic recurrent model for next basket recommendation. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 729–732.
- [9] Guorui Zhou, Xiaoqiang Zhu, Chengu 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*. ACM, 1059–1068.