KRAF: A Flexible Advertising Framework using Knowledge Graph-enriched Multi-Agent Reinforcement Learning

Jose A. Ayala-Romero, Peter Mernyei, Bichen Shi, Diego Mazón
Huawei Ireland Research Center

Motivation

Real-time bidding (RTB) is the process in which digital advertising inventory is bought and sold on a per-impression basis, via instantaneous programmatic auction.

Bidding optimization is one of the most important and challenging problems in online advertising. Most of the advertising platforms offer **auto-bidding tools** to help the advertiser to allocate their budget.

Limitations:

- Rely on click-through rate (CTR) models that use sequential user-ad interactions (user tracking).
- Due to the new privacy regulations (e.g., GDPR), user data is more restricted (e.g., 3rd party cookies are expected to eventually disappear).
- New advertising markets will provide heterogeneous information (e.g., Autonomous advertising).

Motivation (II)

Advertisers need to allocate their budget over time by sequentially selecting a bid value for each ad impression slot. While other works formulate the problem for a single bidder considering the market price stationary, the auction mechanism is actually a multi-agent system by nature.

The outcome of each advertiser is highly dependent on the actions of all the involved bidders.

Some additional aspects that make this problem very challenging:

- User response (clicks, conversions) is very sparse.
- The budget of the advertisers is usually limited and it is desirable to spend it throughout the whole day.
- Prioritizing some advertiser can increase the benefit of the platform in the short term (e.g., advertisers with higher CTR or budget), but it can be disadvantageous in the long term.

Motivation (III)

We aim to design an auto-bidding tool that can efficiently use **heterogeneous data,** adapt to user different user **privacy settings,** and solve the **multi-agent budget allocation problem**.

We propose *Knowledge Graph-enriched Multi-Agent Reinforcement Learning Advertising Framework* (**KRAF**).

The main ideas are summarized as follows:

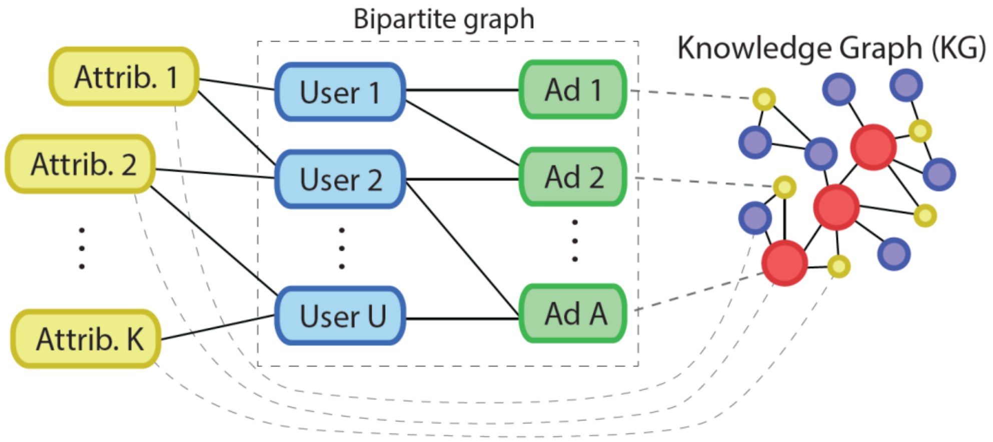
- We represent the input information as a graph and using **Knowledge Graph (KG)** techniques we compute dense representations (embeddings) of the graph nodes.
- The KG allows us to integrate heterogeneous information and adapt to different levels of privacy.
- We design a **Multi-agent Reinforcement Learning (MARL)** algorithm that leverages these embeddings to perform the multi-agent budget allocation.

Problem formulation

We represent user-ad interactions in a bipartite graph. We connect attributes to users and integrate all the entities into the KG. This graph is referred to as Unified Knowledge Graph (UKG).

The UKG captures the high-order connections between ads, users and their attributes.

Note that the specific meaning of the graph entities depends on the privacy setting (e.g., contextual advertising, Interest based advertising, etc.)



Our objective is to distill dense vector representations (embeddings) of users and ads. These embeddings will be used to build the state for the decision-making algorithm in order to solve the multi-agent budget allocation problem.

Moreover, based on these embeddings, we also predict the interest of users in ads by learning the affinity $\delta_{u,z}$ between user u and ad z.

Problem formulation (II)

We formulate the budget allocation in the sequential second price auctions as a stochastic game given by $\Gamma =< N, \mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P}, \gamma >$

We define the policy for an agent i as $\pi^i : \mathcal{S}^i \mapsto \mathcal{A}^i$, an a joint policy is defined as $\pi = (\pi^1 \ldots \pi^N)$

The value function of a certain joint policy π for the agent i is given by

$$v_\pi^i(\mathbf{s}) = \sum_{t=0}^{T} \gamma^t \mathbb{E}_{\pi,\mathcal{P}} \left[r_{i,t}(\mathbf{s}_t, \mathbf{a}_t) | \mathbf{s}_0 = \mathbf{s}, \pi \right]$$

The objective of the agents is to find the optimal policy $\pi*$ that maximizes their value function shown in eq. (1). Note that each player has an independent reward function that depends on the behavior of other agents. Therefore, the definition of the optimal policy is not clear in this setting as the maximization of the value function for one user can harm the value function for others, making the game non-stationary. To solve this problem we propose an equilibrium-based approach [41]:

$$v^i(s; \pi_*) \geq v^i(s; \pi^i, \pi_*^{-i}).$$

where π_*^{-i} denotes the joint policy of all the agents except agent i, i.e., $\pi_*^{-i} = [\pi_*^1, \ldots, \pi_*^{i-1}, \pi_*^{i+1}, \ldots \pi_*^N]$.

KRAF: UKG-Enriched State Representation

To learn the dense representation (embeddings) of the graph entities, we use a hybrid approach that combines Knowledge Graph Embedding (KGE) models and graph convolution networks (GCN).

- The KGE models parameterize entities and relations as vector representations preserving the structural properties of the graph. We apply **TransR**.
- The GCN propagates recursively the information of the nodes. The embedding of each node is updated based on the embeddings of its neighbors, thus capturing high-order connectivity. We apply GCN with attention mechanism or **Graph Attention Networks (GAT)**.

We learn the embeddings that minimize the global loss given by:

$$\mathcal{L} = \mathcal{L}_{\mathrm{KGE}} + \mathcal{L}_{\mathrm{GCN}} + \lambda||\theta||_2^2,$$

The embedding of a user u is given by $e_u^* = concat\{e_u^0, e_u^1, \dots, e_u^L\}$, where e_u^j is the embedding representation at layer j of the GCN.

Finally, the affinity between user u and ad z is given by $\delta_{u,z} = {e_u^*}^\top e_z^*$.

KRAF: Multi-agent Reinforcement Learning for budget allocation

- We assign one MARL agent to each advertiser to make decisions on its behalf.
- The graph embeddings compose the state of the MARL agents.
- Each MARL agent select the bidding value of its associated advertiser for each ad slot.

The observation of the state of agent i at time t is defined as $o_t^i = [b_t^i, \bar{b}_t^i, e_i^*, e_{u_t}^*, \delta_t^i]$

User-ad affinity

Budget of agent i

Expected budget of agent i

Embeddings of user u
and advertiser i

We define the action for agent i at time t as $a_t^i \in [\beta_{min}, \ \beta_{max}]$

The reward for agent i at time t is given by two terms: $r_{i,t} = r_{i,t}^{\text{auct}} + \beta \cdot r_{i,t}^{\text{budget}}$

Reward given to the winner of the auction

$$r_{i,t}^{\text{auct}} = \begin{cases} r_{i,t}^{\text{bid}}, & a_t^i > a_t^j \quad \forall j \neq i \\ 0, & \text{otherwise,} \end{cases}$$

$$r_{i,t}^{\text{bid}} = \begin{cases} 1, & \text{if user-ad interaction from } \mathcal{M}^i \\ \mathcal{F}(\delta_t^i), & \text{otherwise} \end{cases}$$

Rank-based reward to reduce the sparsity

Shaped reward penalty
(encourages the agents spend the full budget throughout the day but not run out of it too early)

$$r_{i,t}^{\text{budget}} = -\left|b_t^i - \bar{b}_t^i\right|$$

KRAF: Large-Scale Multi-Agent System Design

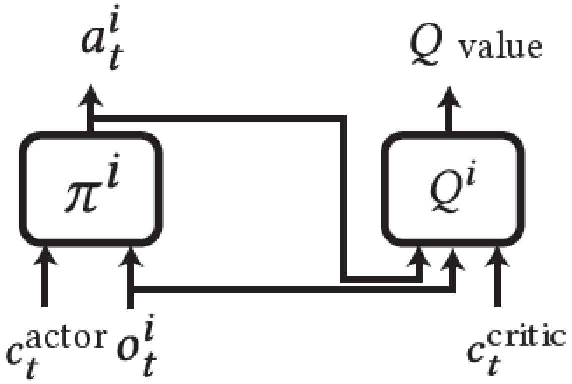
Most MARL methods are limited typically to a small number of agents. When the number of learning agents is very large, the learning process becomes intractable as the computational complexity grows exponentially with the number of agents (curse of dimensionality).

There are a few works addressing the scalability problem by including ideas from Mean Field Theory. However, some of the assumptions made in these previous works do not hold in our problem. First, they consider discrete action spaces, while in our case the bids are continuous. Second, the average effect of the population (e.g., average bid value, average affinity) is not very informative in our case, hindering the coordination among learning agents and convergence.

Based on that, we design two coordination signals based on domain knowledge of this setting. We define the critic and actor coordination signals as:

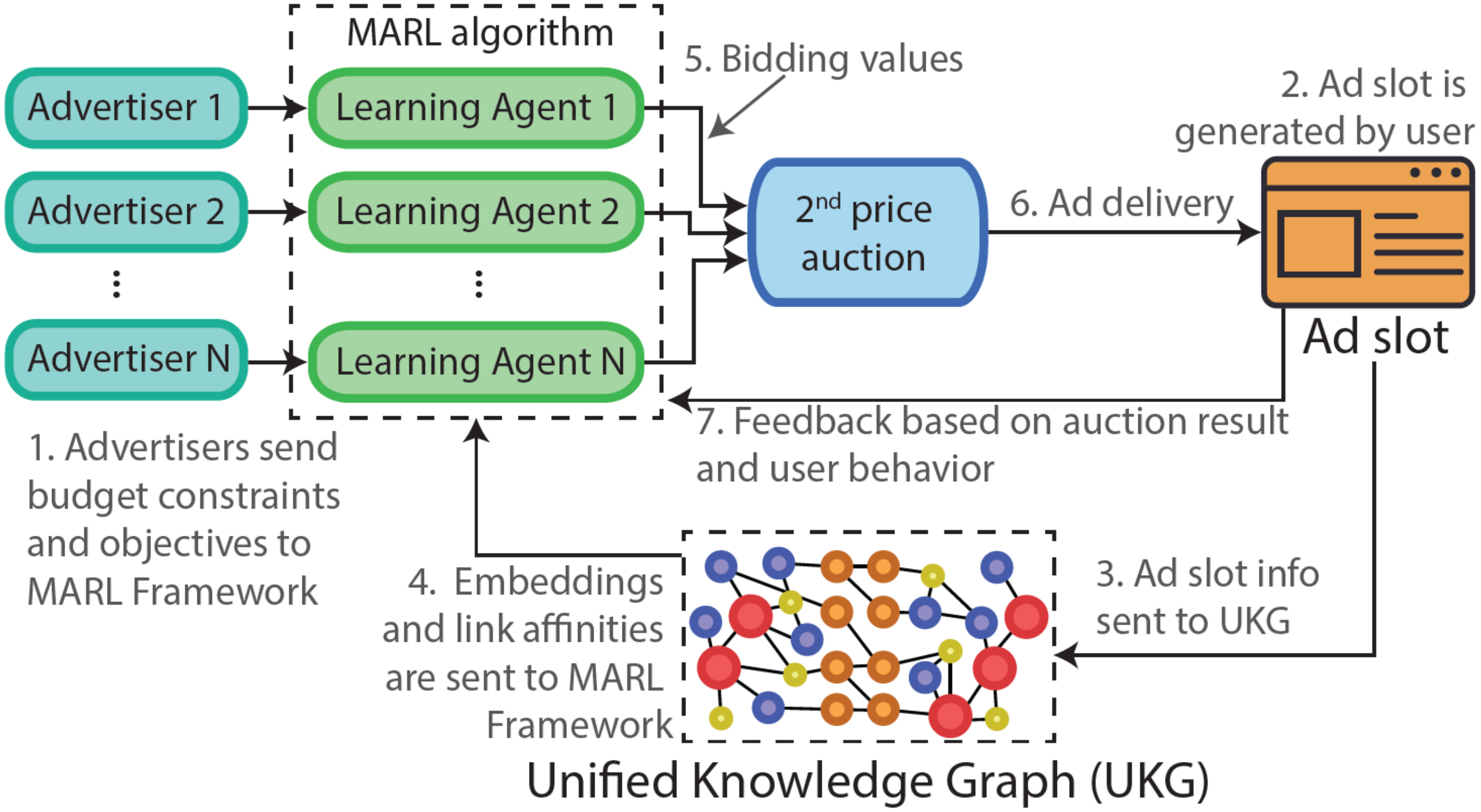
$$c_t^{\text{critic}} = \left(\delta_t^{\max}, \delta_t^*, a_t^{\max}, MP\right), \quad c_t^{\text{actor}} = \left(\delta_t^{\max}\right)$$

These coordination signals encode the state of the game more effectively than the average state or action of the population, avoiding non-stationarity and enabling convergence.



Architecture of the MARL algorithm

KRAF: Architecture and dataflow.


MARL algorithm
Advertiser 1
Advertiser 2
Advertiser N
Learning Agent 1
Learning Agent 2
Learning Agent N
5. Bidding values
2nd price auction
6. Ad delivery
2. Ad slot is generated by user
Ad slot
7. Feedback based on auction result and user behavior
1. Advertisers send budget constraints and objectives to MARL Framework
4. Embeddings and link affinities are sent to MARL Framework
3. Ad slot info sent to UKG
Unified Knowledge Graph (UKG)

Evaluation: Datasets and metrics.

We evaluate our proposal using three datasets: Amazon-book, Last-FM, Yelp2018.

The evaluation is based on four metrics:

- **ROAS (Return On Ad Spend)** measures the revenue over the cost. We are interested in measuring the global ROAS (across all the advertisers) to compare different strategies. We consider that the revenue is equal to 1 when the user clicks and 0 otherwise. Thus, ROAS = (Total # of clicks)/(Total budget spent)

- **# clicks** is the accumulated number of click interactions for all the advertisers during one day.

- **Fairness** measures the similarity on the ROAS across advertises. We use the Jain's fairness index across the individual ROAS of the advertisers:

$$\text{Fairness} = \frac{\left(\sum_{i=1}^{N} \text{ROAS}_i\right)^2}{N \sum_{i=1}^{N} \text{ROAS}_i^2}$$

Evaluation: Benchmarks

- **Random policy.** Advertisers select a random bid in $[\beta_{min}, \beta_{max}]$ at each time t. This policy is used as baseline for comparison.

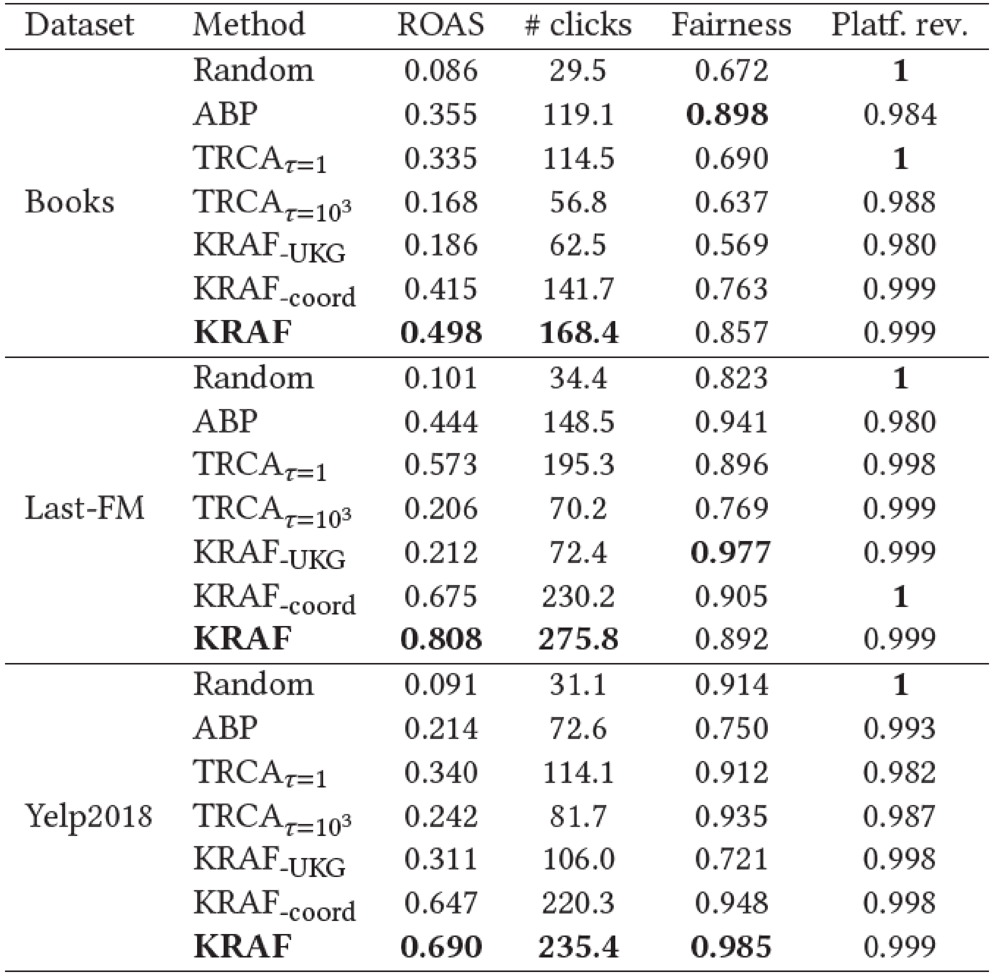
- **Affinity-based Policy (ABP).** A common approach in online advertising is to select the bid for each advertiser proportional to its pCTR. Using affinity as a proxy, the advertisers select a bid proportional to their affinity ($\delta_{u,z}$).

- **TRCA [38].** We adopt a temperature regularized credit assignment (TRCA) to distribute the reward among the agents. This strategy distributes the reward among the agents according to their contribution to the auction using a softmax function with temperature parameter τ.

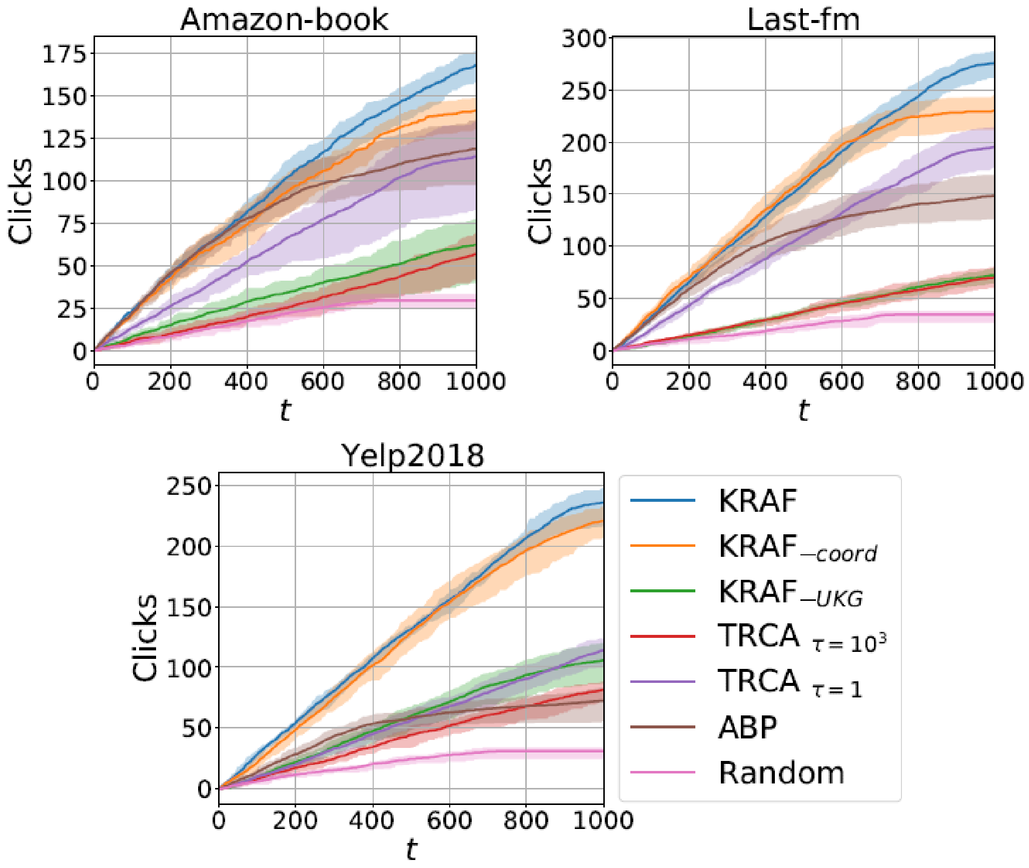
- **KRAF$_{\text{-UKG}}$.** We disable the use of the UKG in our proposal, using only the bipartite graph to compute the embeddings and the link prediction. For that purpose, we compute the embeddings using Neural Graph Collaborative Filtering (NGCF) [35]. Our goal is to evaluate the impact of KG on the final performance.

- **KRAF$_{\text{-coord}}$.** We disable the coordination signals for both the actor and the critic (i.e., c_t^{actor} and c_t^{critic}). The objective is to evaluate their the impact on the final performance.

Evaluation: Results



Dataset	Method	ROAS	# clicks	Fairness	Platf. rev.
Books	Random	0.086	29.5	0.672	**1**
	ABP	0.355	119.1	**0.898**	0.984
	TRCA$_{\tau=1}$	0.335	114.5	0.690	**1**
	TRCA$_{\tau=10^3}$	0.168	56.8	0.637	0.988
	KRAF$_{\text{-UKG}}$	0.186	62.5	0.569	0.980
	KRAF$_{\text{-coord}}$	0.415	141.7	0.763	0.999
	KRAF	**0.498**	**168.4**	0.857	0.999
Last-FM	Random	0.101	34.4	0.823	**1**
	ABP	0.444	148.5	0.941	0.980
	TRCA$_{\tau=1}$	0.573	195.3	0.896	0.998
	TRCA$_{\tau=10^3}$	0.206	70.2	0.769	0.999
	KRAF$_{\text{-UKG}}$	0.212	72.4	**0.977**	0.999
	KRAF$_{\text{-coord}}$	0.675	230.2	0.905	**1**
	KRAF	**0.808**	**275.8**	0.892	0.999
Yelp2018	Random	0.091	31.1	0.914	**1**
	ABP	0.214	72.6	0.750	0.993
	TRCA$_{\tau=1}$	0.340	114.1	0.912	0.982
	TRCA$_{\tau=10^3}$	0.242	81.7	0.935	0.987
	KRAF$_{\text{-UKG}}$	0.311	106.0	0.721	0.998
	KRAF$_{\text{-coord}}$	0.647	220.3	0.948	0.998
	KRAF	**0.690**	**235.4**	**0.985**	0.999



Industrial Use Cases & Ads Privacy

KRAF can be adapted to different settings related to user privacy protection in advertising services:

Contextual advertising. Contextual advertising is an approach that only uses the content the user is consuming when the ad slot is generated to recommend ads.

KRAF can be integrated into contextual advertising to enhance the richness of the information. We represent all the contextual tags as graph entities connected to the user who generated the ad slot. As all these tags have a corresponding entity in the KG, users and ads are indirectly connected through the UKG even when there is no past behavioral information about the users. Thus, via leveraging high-order connectivity, the advertising platform can provide high quality recommendations

Cohort-based advertising. It relies on creating cohorts (or groups) of users with similar interests (e.g., behavioral data). Thus, when generating ad slots, users only reveal their cohort instead of behavioral or sensitive information.

We can employ KRAF by considering cohorts of users instead of individual users in the UKG and record cohort-ads interactions. This will decrease the sparsity of our graph as a cohort will have more interaction (e.g., clicks) than an individual user, and increase the accuracy of predicted affinities. Similarly, we could define richer types of relations, for instance, different degrees of affinity relations based on the number of clicks. Thus, we increase the expressiveness of the model.

Industrial Use Cases & Ads Privacy

Interest-Based Advertising (IBA). At every epoch (e.g., one week) users generate a list of topics based on their behaviors (e.g., website visits or installed apps). A topic is a human-readable tag taken from the topics taxonomy set. When a user generates an ad slot, a subset of the topics is sent for ad selection.

In our framework, each topic can be modeled as a graph entity connected to the UKG. When a user generates an ad slot, her past interactions are not available (the user is anonymous), but she will be connected to her topics and therefore integrated into the UKG. Although storing interaction with ads of individual users is not an option, we can store interactions between anonymous users' topics and ads, and thus, learn the high-order connectivity relations among them.

New Advertising Markets. New advertising markets are emerging in recent years and adaptability is a key factor for advertising platforms. Autonomous Advertising is an example, where companies offer free rides in autonomous cars in exchange for showing ads during the ride.

In that case, assuming that we do not have previous information about the user (e.g., from the app to request the ride), other information can be added to the graph such as the origin and destination of the ride, the time, the day of the week, and so on. These graph entities can improve the targeting of the ads by providing geolocated advertisement adapted to the time and context of the users.

Thank you!

HUAWEI

CIKM

Atlanta, GA 2022