

Automated Creative Optimization for E-Commerce Advertising

Jin Chen*

University of Electronic Science and
Technology of China
chenjin@std.uestc.edu.cn

Ju Xu

Alibaba Group
xu.ju.xj@alibaba-inc.com

Gangwei Jiang*

University of Science and Technology
of China
gwjiang@mail.ustc.edu.cn

Tiezheng Ge, Zhiqiang Zhang

Alibaba Group
{tiezheng.gtz,zhang.zhiqiang}@alibaba-
inc.com

Defu Lian

University of Science and Technology
of China
liandefu@ustc.edu.cn

Kai Zheng

University of Electronic Science and
Technology of China
zhengkai@uestc.edu.cn

ABSTRACT

Advertising creatives are ubiquitous in E-commerce advertisements and aesthetic creatives may improve the click-through rate (CTR) of the products. Nowadays smart advertisement platforms provide the function of compositing creatives based on source materials provided by advertisers. Since a great number of creatives can be generated, it is difficult to accurately predict their CTR given a limited amount of feedback. Factorization machine (FM), which models inner product interaction between features, can be applied for the CTR prediction of creatives. However, interactions between creative elements may be more complex than the inner product, and the FM-estimated CTR may be of high variance due to limited feedback. To address these two issues, we propose an Automated Creative Optimization (AutoCO) framework to model complex interaction between creative elements and to balance between exploration and exploitation. Specifically, motivated by AutoML, we propose one-shot search algorithms for searching effective interaction functions between elements. We then develop stochastic variational inference to estimate the posterior distribution of parameters based on the reparameterization trick, and apply Thompson Sampling for efficiently exploring potentially better creatives. We evaluate the proposed method with both a synthetic dataset and two public datasets. The experimental results show our method can outperform competing baselines with respect to cumulative regret. The online A/B test shows our method leads to a 7% increase in CTR compared to the baseline.

KEYWORDS

Advertising Creatives, Exploration and Exploitation, AutoML, Thompson Sampling, Variational Bayesian

*This work was done when the authors Jin Chen and Gangwei Jiang were at Alibaba Group for intern.

†Corresponding author

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.

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8312-7/21/04.

<https://doi.org/10.1145/3442381.3449909>

ACM Reference Format:

Jin Chen, Ju Xu, Gangwei Jiang, Tiezheng Ge, Zhiqiang Zhang, Defu Lian, and Kai Zheng. 2021. Automated Creative Optimization for E-Commerce Advertising. In *Proceedings of the Web Conference 2021 (WWW '21)*, April 19–23, 2021, Ljubljana, Slovenia. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3442381.3449909>

1 INTRODUCTION

Online advertisements are ubiquitous in nowadays life, creating considerable revenue for many e-commerce companies. As a common medium of advertisements, advertising creatives, as shown in Fig. 1, can deliver rich product information quickly to users in a visual manner. Appealing creatives improve visual experience and may lead to an increase of click-through rate (CTR), as evidenced by [2, 6]. For merchants and e-commerce companies, the increase of CTR can be considered as an indicator of an increase of revenue. Therefore, much attention has been paid to creative design for improving the visual experience.

Traditionally, advertisers have to employ expert designers to produce attractive creatives and then submit them to platforms as complete images. Each time new products are announced or old products are updated, many creatives of different sizes and styles are required to design and submit to advertising platforms. This leads to an expensive cost for many advertisers. To reduce the cost of repetitive but professional design for advertisers, several high-tech companies set up intelligent advertisement platforms [11], which provide instant production services for advertising creatives and remarkably reduces heavy burdens for advertisers. Advertisers only need to provide basic materials to platforms, such as product pictures and textual information. Based on these source materials, the production system produces advertising creatives automatically by compositing arbitrarily designated elements, such as templates, colors of text and sizes of pictures.

In order to ensure the quality of generated creatives, on one hand, they should satisfy basic visual constraints, but this is not the focus of this paper. On the other hand, they should be clicked with high probabilities (i.e., click-through rate) when they are advertised. Intrinsically speaking, the latter corresponds to an optimal selection problem, which faces the following challenges. First, the combinatorial composition of elements leads to an exponential explosion in the number of candidate creatives. Second, because of the limited advertising budget, each product is usually displayed several times within a day. When apportioned to a large number of its generated creatives, the feedback becomes extremely sparse. Furthermore,

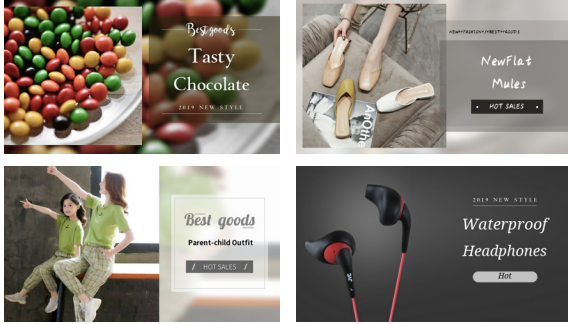


Figure 1: Advertising Creatives with various elements. These creatives are composited with different templates, product images, textual information, text fonts and different background settings.

creatives in E-commerce change frequently over time, so that cumulative feedback for out-of-date products may be not useful any longer. Usually, there are more than 4 million new creatives in a day in a popular advertisement position according to our statistics. Therefore, it is extremely difficult to estimate the click-through rate for each generated creative accurately.

It is possible to apply factorization machines (FM) [31] for predicting the click-through rate (CTR) of each creative. FM models interaction between elements of creative based on inner product, so that creatives with similar composited elements are similarly represented. As a consequence, FM can alleviate the sparsity issue to some extent. However, interactions between elements of creatives may be much more complex than the inner product. For example, we empirically observe that the inner product does not work best for modeling interactions between elements. Moreover, the estimated CTR for each creative may be of high variance due to the extremely sparse feedback. Greedy creative advertising with maximal predicted CTR is usually suboptimal, so that it is essential to efficiently explore potentially better creatives by simultaneously exploiting CTR prediction and uncertainty.

To address these two issues, we propose an Automated Creative Optimization (AutoCO) framework to model complex interaction between elements of creatives and to strike a balance between exploration and exploitation. Particularly, inspired by Automated Machine Learning (AutoML) [12], we propose one-shot search algorithms for searching effective interaction functions between elements of creatives efficiently. The interaction function family to search can be defined by extending the multiply operator in the inner product to the operator set {max, min, plus, multiply, concat} over operation-aware embedding, and replacing the sum pooling operation with a fully connected network. Following that, following the reparameterization trick in VAE [15], we develop stochastic variational inference to estimate the posterior distribution of parameters. Armed with the parameter posterior distribution, we can apply Thompson Sampling [3] for efficiently exploring potentially better creatives.

The contributions in this paper are summarized as follows.

- We propose an Automated Creative Optimization (AutoCO) framework for optimal selection of composited E-commerce

creatives. The framework simultaneously models complex interaction between creative elements and strikes a balance between exploration and exploitation, successfully addressing the sparsity issues of feedback.

- We empirically observe that the inner product is suboptimal for modeling interaction between elements, based on which we propose a one-shot search algorithm for searching effective interaction functions efficiently.
- Based on the reparameterization trick, we develop stochastic variational inference to estimate the posterior distribution of parameters, making it amenable to Thompson Sampling for efficiently exploring potentially better creatives.
- The experiments on both a synthetic dataset and two public datasets show that our method performs better than competing baselines in terms of the cumulative reward and regret, indicating the effectiveness of complex interaction modeling. The online A/B test shows that our method leads to a 7% increase in CTR, confirming the superiority of our method to baselines.

2 RELATED WORK

2.1 Similar Tasks

With the rapid development of the Internet, the recommendation systems have been proposed to solve the problem of information overload, such as online advertising [44], point of interest recommendation [23] and so on. The Creatives are ubiquitous for online advertisements. Some works have paid attention to CTR prediction on ad creatives [5, 26, 29, 39, 42] via extracting expressive visual features to increase the CTR. However, there are few works about the optimization for advertising creatives given limited feedbacks. In classic recommendation systems, the negative samplers [21] are utilized to select informative samplers for solving the data sparsity and the product quantization methods [22] have been used for lightweight recommendation.

For online advertising and recommendation, several similar tasks have been studied. LinUCB [20] achieved great success in personalized news recommendations where an efficient contextual bandit is applied. Whole-page optimization, aiming to select the best layout of a home page, is usually defined as a bandit problem [37]. An efficient multivariate bandit algorithm [10] is proposed to solve the problem and shows superior performance in real online situations. These similar tasks exploit bandit algorithms to solve the exploration and exploitation dilemma.

2.2 Success of AutoML for recommendation

Some success of automated machine learning (AutoML) has been achieved in recommendation systems. Different from regular tasks for AutoML, such as classification in computer vision, search space becomes diverse rather than the neural architectures [43, 45, 47]. Searching embeddings of the input with categorical features is exploited for large scale recommendations [13]. Another attention is that the search for high order features helps improve the CTR prediction [24]. Furthermore, the search for interaction functions for collaborative filtering reveals the improvements caused by searching operation functions between different user embeddings and item embeddings [40].

3 PRELIMINARIES

Before we introduce the proposal in detail, we give a brief description of the investigated problem and introduce the classical method. Finally, we provide an overview of our solution.

In this work, only composited creatives are investigated instead of the whole images designed by professional designers. At the time of impression, a product sends a request to the advertisement platform, and the platform instantly selects the creative from candidates for display. There are M elements to be composited and the i -th element is a collection of n_i alternatives.

The investigated task in this work is to recommend the optimal composited creative for each product at the time of impression, aiming to increase the overall CTR given limited impressions. Noticing that the generation of the creatives under aesthetic metrics, such as [18, 36] is not the focus of this paper.

3.1 Classical Method

Typically, each candidate creative should be delivered to the advertising platforms and the user feedback is collected for CTR prediction with point-wise [5], pair-wise [9] methods. The creative with a maximal value of predicted CTR is then selected to be displayed to increase the overall click-through rate. But the multiple elements are composited combinatorially, resulting in an exponential explosion in the number of creatives. For example, given 4 templates, 10 fonts, 10 colors, and 5 picture sizes, 2,000 creatives can be composited for one of the product images. The collection of sufficient samples for model training is time-consuming and expensive. Thus there is a trade-off between exploration and exploitation.

Multi-Armed Bandit is usually a classical option for creative optimization where the composited creatives are compared to bandits. These algorithms generally include the following processes. At a discrete-time $t \in [1, \dots, T]$, a creative c is selected and revealed with a reward $R_{c_t} \in \{0, 1\}$. An impression is a scenario when an ad is shown to a user in the advertising industry and $R = 1$ represents the user clicked the selected creative. Within the limited impressions, the multi-armed bandit aims to minimize the cumulative regret $\sum_{t=1}^T (R_{c^*} - R_{c_t})$, where $c^* = \arg \max \mathbb{E}(R_c)$ represents the optimal creative with the maximal expected reward. In this work, the expected reward corresponds to the value of CTR.

The composited creatives sometimes share common elements, so that making connections can help estimate the expected reward of various creatives. The linear function to model the connections has been successfully applied in this situation [7], where the expected reward of the creative is assumed as a linear function with respect to the feature. The expected reward for each creative is additionally measured by a upper confidence bound. The linear models ensure the theoretical lower bound for regret. However, the expressiveness of linear models is limited. Such models can not efficiently capture the commonality given significantly large search space. This limitation motivates us to design a more effective algorithm under the huge creative space given extremely sparse feedback.

3.2 Solutions Overview

There are two key components in a bandit problem: (1) Assumption of the expected reward (2) Exploration methodology. For these two parts, we design effective algorithms to improve the performance in

terms of overall CTR under the huge creative space. The proposed framework consists of the following phases:

- (1) **CTR Estimation**: We focus on the interactions between different elements to leverage commonality between numerous creatives, and search for the interaction functions to capture suitable representations depending on AutoML methods.
- (2) **Efficient Exploration**: To reduce the high variance caused by sparse feedback, Thompson Sampling, a classical and efficient exploration method, is exploited. The posterior approximation under complex interaction functions is estimated through variational inference in an end-to-end manner.

4 INTERACTION FUNCTION SEARCH FOR CTR ESTIMATION

4.1 Interaction Function Analysis

In a similar task, the optimal selection of layouts for web pages [10], the interactions between different elements are captured as the way in the Poly2 [4] model. But the parameter space becomes seriously large with the increasing number of features. One solution is the factorization machines (FMs) [31], which utilize the inner product as the interaction functions and achieve success in CTR prediction [8]. The inner product of the pairwise feature interactions is as:

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \sum_{k=1}^d e_{ik} \cdot e_{jk}$$

where $\mathbf{e}_i \in \mathbb{R}^d$ is the low-dimension embedding vector for the i -th feature. $\mathbf{e}_i = V_i \mathbf{x}_i$. $V_i \in \mathbb{R}^{d \times l_i}$ is the embedding matrix with respect to the i -th feature field and $\mathbf{x}_i \in \mathbb{R}^{l_i}$ denotes the one-hot encoding vector of the i -th feature. Multi-hot vectors or continuous value vectors are also available. d is the embedding dimension and l_i represents the size of feature space.

However, the inner product may not yield the best performance in many recommendation tasks due to the complex nature of interactions. Inspired by the recent work SIF [40], which searches for the interaction functions between the user and item embeddings, we are encouraged to capture the complicated interactions between different features. For the sake of simplicity, we state the problem of searching for the interaction functions between different feature fields in this investigated task.

Definition 1 (Interaction Functions Search). Find an interaction function $o_{ij} \in \mathcal{O}$ for each pair of embedding vectors $(\mathbf{e}_i, \mathbf{e}_j)$. The object is to minimize the loss over the observed data.

\mathcal{O} is the operation set containing several interaction functions. We utilize $|\mathcal{O}|$ to represent the number of operations in \mathcal{O} . There are L embedding vectors and the number of the interaction functions to be selected is $K = C_L^2$. Thus the time complexity for searching the optimal interaction function is $\mathcal{O}(|\mathcal{O}|^K)$. In the following chapters, we will introduce the search space and the search algorithm respectively.

4.2 Search Space

We pick the following simple and popular operations as mentioned in the previous work [40]. (1) *Concat*: $[\mathbf{p}; \mathbf{q}]$, (2) *Multiply*: $\mathbf{p} \circ \mathbf{q}$, (3) *Plus*: $\mathbf{p} + \mathbf{q}$, (4) *Max*: $\max(\mathbf{p}, \mathbf{q})$, (5) *Min*: $\min(\mathbf{p}, \mathbf{q})$. The concat is a

Table 1: AUC and the Relative improvements on different interaction functions compared with FM

	Multiply	Concat	Plus	Max	Min	LR	IFS(Alg.1)	FM
Average AUC	0.6039	0.6044	0.6037	0.6039	0.6023	0.5285	0.6050	0.5947
Standard Error	0.0017	0.0004	0.0007	0.0013	0.0008	0.0019	0.0003	0.0004
Relative Improvements	1.54%	1.62%	1.50%	1.53%	1.27%	-11.13%	1.71%	-

vector-wise operation and the other four operations are element-wise. Similar with SIF, we exploit a full connected layer for each function. The FC controls the output size of different operations to be consistent.

Before we focus on the approach of the efficient search for interaction functions, several offline experiments are conducted to demonstrate that different interaction functions between feature fields yield different performances. The baseline approaches involve the factorization machine and the logistics regression without interactions.

As shown in Table 1, different operations have been studied on the collection of about one thousand products and their 300 thousand composited creatives. Nearly three million pieces of data are collected. We use the regular CTR prediction with the binary cross-entropy loss for optimization. The five selected operation functions have better performances than the FM models and have different degrees of improvement. The result provides evidence that there exist more approximate interaction functions between the different elements.

Considering all the interactions between different embedding vectors, the search space is up to 5^K and is much more complicated. The various operators between different embedding vectors may enhance the expressiveness of the CTR predictor and then lead to better performances than the FM-based models.

4.3 Efficient Search Strategy

A straightforward idea to obtain the optimal interactions between different feature fields is to traverse all the combinations. However, the time complexity of collecting the best interactions is $O(|O|^K)$, which is NP-hard and unacceptable in our situation. To deal with the issue, we apply an efficient search algorithm inspired by previous AutoML algorithms.

4.3.1 One-Short Search Algorithm. Neural architecture search is firstly proposed in [46], where reinforcement learning is employed for searching architectures. After then, more searched based algorithms have been proposed (e.g., evolution algorithm [27], greedy search [25]). However, these algorithms cost much search time due to the huge and discrete search space. Recently, the one-shot architecture search algorithms achieve huge success depending on the continuous search space and greatly shorten the time, such as DARTS [28], SNAS [38], NASP [41]. These algorithms jointly optimize the weights for different operators and model parameters by stochastic gradient descent.

Following SIF [40], we relax the choices among operations as a sparse vector in a continuous space. For each pair of feature fields i and j , $\alpha_{ij} \in \mathbb{R}^{|O|}$ denotes the weight vector and the k -th element $\alpha_{ij}^{(k)}$ corresponds to the weight of the k -th operator $O_{ij}^{(k)}$. Now the

Algorithm 1: Interaction Function Search (IFS)

Input: Observations \mathcal{D} , learning rate η

- 1 **while** *Not Converge* **do**
- 2 Select the interaction function $\tilde{\alpha}_{ij} = \text{prox}_{C_1}(\alpha_{ij})$ for each pair;
- 3 Calculate $F(\Theta, \tilde{\alpha})$ according to Eq (1) and (2);
- 4 Update continuous α of operation weights
 $\alpha = \text{prox}_{C_2}(\alpha - \eta \nabla_{\tilde{\alpha}} F(\Theta, \tilde{\alpha}))$;
- 5 **end**
- 6 **return** *Interaction functions between each pair of embedding vectors parameterized by α*

operation \tilde{O}_{ij} between the feature pair i and j is formulated as follows:

$$\tilde{O}_{ij}(X) = \sum_{k=1}^{|O|} \alpha_{ij}^{(k)} O_{ij}^{(k)}(X), \text{ where } \alpha_{ij} \in C_1 \cap C_2.$$

$$C_1 = \{\alpha \mid \|\alpha\|_0 = 1\}, \quad C_2 = \{\alpha \mid 0 < \alpha^{(k)} < 1\}$$

The constraint C_1 ensures that only one operator would be selected, and the constraint C_2 makes the algorithm more stable. Thus, the search problem is defined as:

$$\begin{aligned} \min_{\Theta, \alpha} \quad & F(\Theta, \alpha) = \sum_{(x, y) \in \mathcal{D}} \ell(I[x, \alpha, \Theta], y) \\ \text{s.t.} \quad & \alpha = \{\alpha_{ij} \mid i = 1, \dots, L, j = i + 1, \dots, L\} \\ & \alpha_{ij} \in C_1 \cap C_2 \end{aligned} \quad (1)$$

where $I[x, \alpha, \Theta] = \sum_i^L \sum_{j=i+1}^L \sum_{k=1}^{|O|} \alpha_{ij}^{(k)} O_{ij}^{(k)}(e_i, e_j)$, α denotes the operation weights for all the $|O|^K$ interactions. ℓ is the binary cross-entropy loss in this work. Θ represents the parameters for embedding matrices and the fully connected layers. For the sake of simplicity, we temporarily omit the expression of the first-order term and the bias in FMs models. Following previous successful work with proximal steps [40, 41], we utilize the same optimization method for efficient searching.

4.3.2 Operation-aware Embedding. An interesting point proposed in FFM [14] and FwFM [30], that features from one field often interact differently with that from other fields, prompts us the potential distinct influence of operators on feature embeddings. Noticing that the joint optimization for the operation weights and embedding vectors, we design the operation-aware Embedding module, shorted as OAE, which assigns different embeddings for each operator. Fig. 2 represents the interactions of different operations.

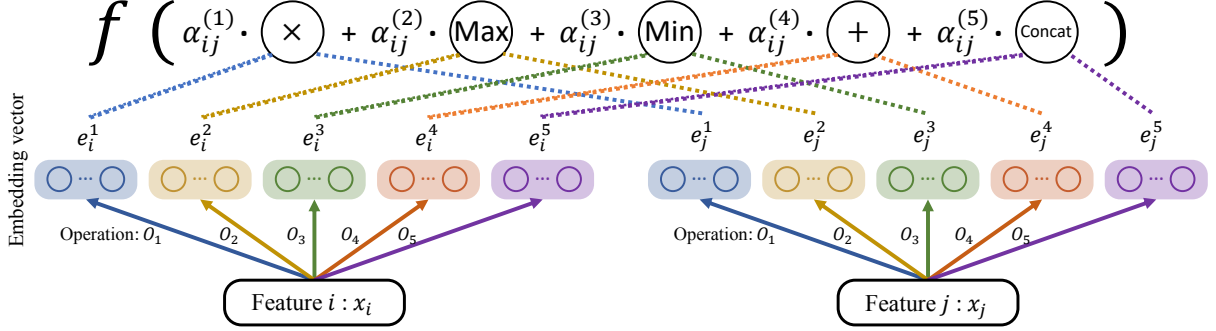


Figure 2: Representation of the search space with the Operation-Aware Embedding module.

According to the operation-aware Embedding, \mathcal{I} is redefined as:

$$\mathcal{I}[\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\Theta}] = \sum_{i=1}^L \sum_{j=i+1}^L \sum_{k=1}^{|O|} \alpha_{ij}^{(k)} O_{ij}^{(k)}(e_i^{(k)}, e_j^{(k)}) \quad (2)$$

The embedding $e_i^{(k)}$ represents the embedding vector corresponding to the k -th operation. Observing the original Eq (1), the shared embedding matrices would obtain gradients of different magnitudes with the change of the operator. A comprehensible example is a difference between the plus and multiply operations. The frequent change would affect the computation of the gradient and result in poor convergence. The introduction of the operation-aware embedding removes the interferences of other operations when optimizing the embeddings and may lead to better learning. The final algorithm for interaction function search with the OAE module is described in Alg. 1.

After the model training, the interaction function O_{ij} is selected by $O_{ij} = \arg \max \alpha_{ij}$. The expected reward of the candidate creative c is followed as:

$$\mathbb{E}[r|\mathbf{x}_c] = \sigma \left(\sum_{i=1}^L \sum_{j=i+1}^L O_{ij} (e_i^{O_{ij}}, e_j^{O_{ij}}) \right) \quad (3)$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function. Then, the creative with the highest expected reward will be selected and revealed with a reward.

5 EFFICIENT EXPLORATION VIA THOMPSON SAMPLING

As mentioned before, the platform can composite hundreds of creatives. But the limited impressions of the products result in the extreme sparsity of feedback. Each candidate creative would be displayed with few impressions so that the CTR predictor comes with high variance. Thus, an efficient exploration method is introduced to reduce the high variance.

5.1 Variational Inference for Thompson Sampling

Thompson Sampling [35] is a popular bandit algorithm to balance exploration and exploitation. The main idea is to assume a simple prior distribution of the reward distribution of each candidate, and

at any time $t \in [1, \dots, T]$, recommend a candidate proportionally to the posterior distribution of being optimal. As declared in [1], instead of solving the posterior distribution of the reward distribution, Thompson Sampling for contextual bandits samples the model parameters. A general process usually includes the following steps:

- (1) Sample model weights from the posterior $\tilde{\boldsymbol{\Theta}} \sim p(\boldsymbol{\Theta}|\mathcal{D})$;
- (2) Recommend the creative $c = \arg \max P(r|\mathbf{x}, \tilde{\boldsymbol{\Theta}})$;
- (3) Receive reward r ;
- (4) Update past observations \mathcal{D} and the posterior $p(\boldsymbol{\Theta}|\mathcal{D})$.

For linear models, given the Gaussian distribution as priors, the posterior has closed-form solutions following Bayesian rules [1, 19]. However, by introducing the complex operations between different feature fields, the posterior distribution is intractable to be directly solved. Thanks to the previous success of posterior approximation in neural networks [3, 15], we can exploit variational inference for approximation to fit in the Thompson Sampling method.

Assuming that the model parameters satisfy the Gaussian distributions, variational learning finds the parameters μ, Σ of the approximation posterior $q(\boldsymbol{\Theta}; \mu, \Sigma) = \mathcal{N}(\mu, \Sigma)$ by minimizing the Kullback-Leibler (KL) divergence between the approximation posterior and the truth posterior:

$$D_{KL}(q(\boldsymbol{\Theta}; \mu, \Sigma) || p(\boldsymbol{\Theta}|\mathcal{D})) = E_{\mathcal{N}(\boldsymbol{\Theta}; \mu, \Sigma)} \left[\log \frac{\mathcal{N}(\boldsymbol{\Theta}; \mu, \Sigma)}{p(\boldsymbol{\Theta}|\mathcal{D})} \right]$$

where \mathcal{D} is the set of observed data and each sample is the feature vector \mathbf{x} and corresponding reward y . Following Bayesian rules, the KL-divergence can be rewritten as:

$$E_{\mathcal{N}(\boldsymbol{\Theta}; \mu, \Sigma)} \left[\log \frac{\mathcal{N}(\boldsymbol{\Theta}; \mu, \Sigma)}{p(\boldsymbol{\Theta})p(\mathcal{D}|\boldsymbol{\Theta})} \right] + \log p(\mathcal{D})$$

The component $\log p(\mathcal{D})$ is independent of the learning parameters μ and Σ , so that the object equals to minimize the following equation:

$$\begin{aligned} \mathcal{J} &= E_{\mathcal{N}(\boldsymbol{\Theta}; \mu, \Sigma)} [\log \mathcal{N}(\boldsymbol{\Theta}; \mu, \Sigma) - \log p(\boldsymbol{\Theta}) - \log p(\mathcal{D}|\boldsymbol{\Theta})] \\ &= D_{KL}(\mathcal{N}(\boldsymbol{\Theta}; \mu, \Sigma) || p(\boldsymbol{\Theta})) - E_{\mathcal{N}(\boldsymbol{\Theta}; \mu, \Sigma)} [\log p(\mathcal{D}|\boldsymbol{\Theta})] \end{aligned}$$

According to Equation (3), the likelihood can be formulated as:

$$p(\mathcal{D}|\boldsymbol{\Theta}) = \prod_{(\mathbf{x}, y) \in \mathcal{D}} (\sigma(\mathcal{I}[\mathbf{x}]))^y (1 - \sigma(\mathcal{I}[\mathbf{x}]))^{1-y}$$

Algorithm 2: Interaction Function Search with Thompson Sampling

Input: Observations \mathcal{D} , learning rate η

- 1 Initialize the parameters μ and Σ ;
- 2 **while** *Not Converge* **do**
- 3 Select the interaction function $\tilde{\alpha}_{ij} = \text{prox}_{C_1}(\alpha_{ij})$ for each pair;
- 4 Sample noise $\epsilon \sim \mathcal{N}(0, 1)$;
- 5 Reparameterize embedding parameters $\tilde{\Theta} = \mu + \Sigma^{\frac{1}{2}} \circ \epsilon$;
- 6 Calculate $F(\tilde{\Theta}, \tilde{\alpha})$ according to Eq (1)(2);
- 7 Update continuous α of operation weights
 $\alpha = \text{prox}_{C_2}(\alpha - \eta \nabla_{\tilde{\alpha}} F(\tilde{\Theta}, \tilde{\alpha}))$;
- 8 Update μ, Σ of variational distribution of embedding matrices corresponding to the selected operation;
- 9 **end**
- 10 **return** *Interaction functions parameterized by α , Model parameters parameterized by μ, Σ*

The logarithm can be simplified as:

$$\begin{aligned} \log p(\mathcal{D}|\Theta) &= \sum_{(x,y) \in \mathcal{D}} y \log \sigma(I[x]) + (1-y) \log(1 - \sigma(I[x])) \\ &= \sum_{(x,y) \in \mathcal{D}} \ell[x, y|\Theta] \end{aligned}$$

Usually, the stochastic variational inference is applied to estimate $E_{\mathcal{N}(\Theta;\mu,\Sigma)}[\log p(\mathcal{D}|\Theta)]$. Suppose that the variational posterior is a diagonal Gaussian distribution, we reparameterize the model parameters which yields a posterior sample of the parameters as $\Theta^{(s)} = \mu + \Sigma^{\frac{1}{2}} \circ \epsilon$. The noises $\epsilon \sim \mathcal{N}(0, 1)$ has the same size as Θ and \circ is point-wise multiplication. Finally, the objective function is formulated as:

$$\mathcal{L}(\Theta^{(s)}; \mu, \Sigma) = D_{KL}(\mathcal{N}(\Theta; \mu, \Sigma) || p(\Theta)) - \sum_{(x,y) \in \mathcal{D}} \ell[x, y|\Theta^{(s)}] \quad (4)$$

where the prior $p(\Theta)$ is assumed as Gaussian distributions in this work. The KL divergence of two Gaussian distributions has an analytical expression. Thus, stochastic gradient descent can be obtained to minimize the objective function. At each iteration, we draw same Gaussian noise ϵ and evaluate Equation (4) and its derivations w.r.t. μ, Σ . The approximation posterior is then estimated as the parameter updates within training iterations.

Integrated with the interaction function search, the whole optimization procedure of Thompson Sampling is declared in Alg. 2. The proposed approach is updated in an end-to-end manner without multiple repetitions of running after the interaction functions update. As mentioned in [3, 17], Non-Bayesian approaches, which trains multiple models for estimating the expected reward, is prohibitively expensive in practice. The application of variational inference facilitates a quick response in a live environment.

5.2 Automated Creative Optimization for E-Commerce Advertisement

Following the above Thompson Sampling method, we can sample a new set of model parameters to estimate the expected reward and pick the creative with the highest value for display. The creative is

Algorithm 3: Automated Creative Optimization for E-Commerce Advertisement (AutoCO)

Input: Feature vectors X of the products and candidate creatives

- 1 Initialize μ and Σ randomly for embedding matrices;
- 2 Initialize the interaction functions parameterized by α ;
- 3 Observations $\mathcal{D} = \emptyset$;
- 4 **for** $t=1,2,\dots,T$ **do**
- 5 Receive the request of product p ;
- 6 Select the interaction functions $O_{ij} = \arg \max \alpha_{ij}$ for each pair of interactions;
- 7 Obtain the embedding vectors \tilde{e} after sampling parameters $\Theta \sim \mathcal{N}(\mu, \Sigma)$;
- 8 The candidate set C corresponding to the product p ;
- 9 Choose the candidate $c_t = \arg \max_{c \in C} \mathbb{E}[r|x_p, c]$ according to Eq (3);
- 10 Receive a binary reward $R_t = \{0, 1\}$;
- 11 Add the new observation (x_{p,c_t}, R_t) into \mathcal{D} ;
- 12 Update μ, Σ, α according to Alg. 2;
- 13 **end**

then delivered with a reward. After getting the new observations, we search for the interaction functions and embedding matrices simultaneously, as shown in Alg. 2. This end-to-end optimization facilitates efficient search and efficient exploration for creative optimization. The final framework of automated creative optimization is shown in Alg. 3. The algorithm would converge as the observations accumulate.

To this end, we perform an automated interaction function search improving the expressiveness of the model and utilize the Thompson Sampling via variational inference for efficient exploration to reduce the high variance caused by limited feedback.

6 EXPERIMENTS

In this section, experiments¹ are conducted on the synthetic dataset and public datasets to answer the following questions:

- **Q1:** How the automated interaction function search improve the performance compared with the FM model?
- **Q2:** Can the proposed Thompson Sampling with variational inference for posterior approximation explore efficiently with limited feedback?

To further examine the performance of our proposed approach, we conduct the experiment on a live production system.

6.1 Synthetic Results

6.1.1 Simulation Data. Following the synthetic experiments in the predictor work [10], we produce the synthetic data depending on the assumed CTR predictor and simulate the reward for the proceeding of bandit problems. There are five elements for compositing creatives, including template, picture size, text font, background blur, and background color. We generate the expected reward over

¹<https://github.com/alimama-creative/AutoCO>

167 products which uses the representation in Equation 3. On average, each product has 67 composited creatives satisfying visual constraints. The operators for different embedding vectors are initialized randomly and the parameters of embedding matrices are sampled from Gaussian distributions. Finally, each creative has an expected reward (CTR), which is the success probability in the Bernoulli trials to simulate the user feedback. The average value of the generated CTR is 0.0885.

At each trial, we sample a product and the algorithm recommends the creative which owns the estimated maximal CTR. Every 10,000 trials are collected for updating and training to simulate the delayed feedback in real production systems. Each algorithm is run multiple times with 20 batches to reduce the bias caused by randomness. The empirical studies are running in a Linux operating system with the Tesla P100 GPU for accelerating.

6.1.2 Baselines. We choose the following typical algorithms for comparisons where parameters are tuned with the best performance in terms of the overall cumulative CTR.

- **Uniform:** The creative is selected uniformly at random.
- **Egreedy:** Recommend a creative at random with probability 0.1 and recommend the creative with the best statistical reward with probability 0.9.
- **LinUCB** [20]: Each candidate creative has the corresponding linear models with the observed contextual information. LinUCB recommends the creative according to the estimated reward with a upper confidence bound.
- **LinTS:** Similarly with LinUCB, the interactions are constructed by linear models. LinTS utilizes Thompson Sampling for exploration depending on the Bayesian linear regression.
- **MVT** [10]: MVT captures the interactions with a linear model and utilizes Thompson Sampling proposed in this work for exploration.

We add another two **FM**-based algorithms **FM** and **FM-TS**. **FM** recommends the creative which owns the maximal estimated CTR without exploration while **FM-TS** exploits Thompson Sampling as declared in Section 5.1.

Our proposed method **AutoCO** automatically searches interaction functions between different feature fields and utilizes the same exploration method as FM-TS.

6.1.3 Metrics. Generally, bandit algorithms are evaluated in terms of cumulative regret [34] which represents the cost before finding the optimum. In practice, the regret is the difference between the expected reward of the optimal strategy and the empirical performance of the conducted algorithm. The lower regret is better. The value of the cumulative reward is complementary to the cumulative regret and higher is better. We also report the normalized cumulative reward, which is actually the CTR in online advertisements.

6.1.4 Learning Effectiveness. We perform 20 batches, i.e. 2 million impressions to test each algorithm. We report the average performance of 5 repetitions experiments in Fig. 3. There are several findings in our empirical studies.

Finding 1: The interactions between different elements enrich the commonality leverage and lead to better performances for CTR prediction. The linear models (e.g., LinUCB and LinTS) which only build the linear relationship between elements converge to worse

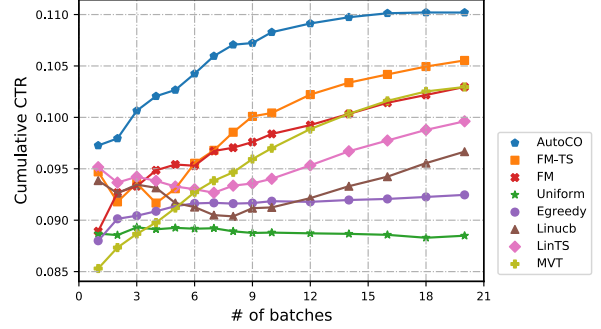


Figure 3: Result of CTR on the simulated data compared with baseline algorithms. Experiments are run for 5 repetitions.

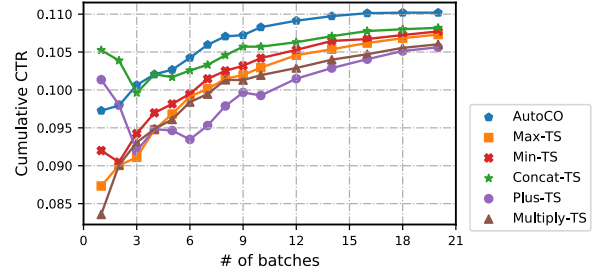


Figure 4: Comparisons of CTR on the different interaction functions.

creatives although they have comparable performances at the beginning. The expressive representations for the interactions help model the commonality of numerous creatives and improve the performances of the overall CTR.

Finding 2: The proposed AutoCO method shows superior performances during the whole process. The significant improvement comes from two key points: (1) Automated interaction function search; (2) Efficient exploration with variation based Thompson Sampling. To investigate the different influences of these two parts, we conduct two experiments.

First, we compare the AutoCO with the fixed single operation in the search space of AutoCO. Results are shown in Fig. 4. All the algorithms are conducted with Thompson Sampling as the exploration method. Among the five operation functions, the concat operation has the best performance than the other four fixed operations while the AutoTS outperforms the concat interaction function. It can be inferred that better interaction functions for different feature fields are found out via the AutoML method.

Second, different exploration methods are tested to compare the efficiency with the Thompson sampling method utilized in this paper. For the FM model and AutoCO, experiments are run for five times to report the average performance in Table 2. The variation based Thompson Sampling achieves excellent performances compared with Greedy and Egreedy. The trivial exploration method,

Table 2: Comparisons of the overall CTR on different exploration methods. The relative improvements of the method exploration are marked in brackets.

	FM	AutoCO*
Greedy	0.1030	0.1073
Egreedy($\epsilon = 0.1$)	0.1024 (-0.58%)	0.1086 (+1.21%)
Thompson Sampling	0.1055 (+2.43%)	0.1102 (+2.70%)

* AutoCO here denotes the interaction function search method for simplicity of understanding.

Egreedy, would always recommend a predefined percentage (ϵ) of creatives randomly although a large number of impressions are collected, causing a waste for displaying bad creatives. Instead, Thompson Sampling recommends the creative proportionally to the probability of being optimum and finally only the optimal creative would be displayed. Another point is that the variational inference for posterior approximation is well conducted in Thompson sampling and this method enables straightforward training in a variety of basic models.

In conclusion, the proposed AutoCO shows the superiority in the accuracy of the CTR predictor and efficiency for exploration on the synthetic data where better expressiveness is obtained through interaction function search.

6.1.5 Effectiveness of OAE. We conduct an ablation study on the simulated data to analyze the effect of the operation-aware embedding module. We first test the AutoCO and then remove the OAE module. The result is reported in Table 3. The OAE module leads to a slight improvement in terms of CTR compared with the shared embeddings between different operations. This indicates the assistance of the independent embeddings. In the process of training, we also find that OAE contributes to more stable learning.

Table 3: Effectiveness of the operation-aware embedding (OAE) module. The experiments are run for 5 repetitions on the simulated data.

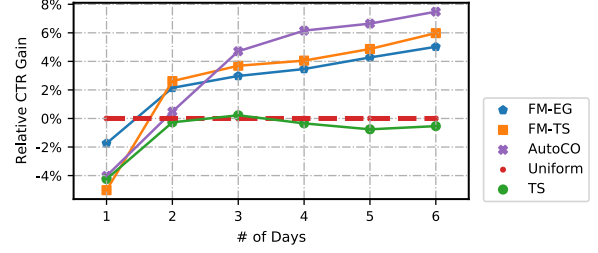
Framework	Overall CTR
AutoCO w/o OAE	0.1097 \pm 0.0006
AutoCO	0.1102 \pm 0.0014

6.2 Experiments on Public Datasets

Empirical studies are also conducted on the two public datasets to verify the performances of our proposed AutoCO.

6.2.1 Datasets.

- **Mushroom:** The Mushroom dataset [33] contains 22 attributes and 8124 samples. The mushrooms are divided into two categories: safe and poisonous. We following the previous settings [3, 32] to formulate the bandit problem, where the algorithm needs to decide whether to eat the given mushroom. If eating a safe mushroom, a reward +5 will be received. Eating a poisonous mushroom obtains the reward +5 with a

**Figure 5: Cumulative CTR of online experiments. We report the relative improvements compared with the Uniform algorithm. There are more than 150,000 impressions per algorithm per day, covering more than 290,000 creatives.**

probability of 0.5 and the reward -35 otherwise. If the agent does not eat the mushroom, no reward is delivered. We sample 50,000 pieces of data from the original dataset.

- **Adult:** The Adult dataset [16] involves 14 attributes of personal information. The task is to determine if a person makes over \$50K a year or not. If the agent makes a right prediction, a reward 1 is delivered. Otherwise, no reward is given. This dataset contains 48842 samples. We pick the 9 categorical features of the mixed 14 original features to conduct our experiments.

6.2.2 Effectiveness of interaction function search. Compared with the single interaction function, the AutoCO has the lowest cumulative regret on the two public datasets, as shown in Table 4. All the algorithms utilize the Thompson Sampling for exploration. The good performance of the AutoCO indicates that the complex interaction functions rather than the single operation enhance the performance of modeling relationships between the different attributes.

6.2.3 Efficiency of exploration methods. To verify the efficiency of Thompson sampling via variational inference, we conduct experiments to compare it with greedy strategy. Experiment results are shown in Table 5. We can conclude that Thompson sampling via variational inference can indeed increase the performance.

6.3 Online Experiment

Now, we examine the performance of our proposed algorithm on the online system.

6.3.1 Experimental Design. We exploit AutoCO for the recommendation of the composited creatives in an online advertisement platform. The composited elements include the template, picture size, text font, background blur, and background color. Our experiments are conducted on one display advertising position of a famous e-commerce company. We generate more than 290,000 creatives satisfying visual constraints covering 1,000 popular products. In addition to the element feature mentioned before, we utilize the category feature of products as well as the contextual information of product images. The feature embeddings of images are extracted from ResNet.

Table 4: Comparisons with single interaction function on two public datasets. Results are relative to the cumulative regret to the Uniform algorithm. Three trials are conducted and we report the mean and standard error. Lower is better.

Dataset	AutoCO	Multiply-TS	Concat-TS	Plus-TS	Min-TS	MAX-TS	Uniform
Mushroom	3.02±0.51	5.44±1.05	6.12±2.37	3.40±0.30	3.62±0.35	4.95±2.00	100.00±0.38
Adult	35.40±0.03	49.28±0.36	35.79±0.06	37.19±0.39	38.34±2.01	36.55±1.18	100.00±0.15

Table 5: Comparisons for different exploration methods on public datasets Mushroom and Adult.

Dataset	AutoCO	AutoCO w/o TS	FM	FM-TS	Egreedy	Uniform
Mushroom	3.02±0.51	5.58±0.51	7.70±1.44	5.64±1.70	63.38±0.47	100.00±0.38
Adult	35.40±0.03	40.07±0.40	43.62±3.32	38.71±2.30	58.85±0.15	100.00±0.15

We performed the experiments for six consecutive days. Traffic for this experiment was randomly and equally separated to the algorithms. The experiments were conducted on non-holidays to avoid fluctuations in traffic. Each algorithm got more than 150,000 impressions per day. The model was updated every hour using the data collection from the beginning to the latest hour.

Baseline algorithms: (1) *Uniform*: The composited creatives are delivered to the platform randomly; (2) *Context-free Thompson Sampling*: The distribution of expected reward is assumed as a Beta distribution. The initialized $\alpha = 1$ and $\beta = 50$; (3) *FM-Egreedy*: FM is the CTR predictor and $\epsilon = 0.2$; (4) *FM-TS*: Apply the proposed Thompson method for exploration in the FM model. In our offline experiments, we found that the interaction functions do not change frequently after several epochs, which is also declared in [41]. In order to reduce the burden of online updating, we do not frequently search the interaction functions.

6.3.2 Result. The six days online experiment shows the superiority of our approach AutoCO, as shown in Fig. 5. The context-free TS algorithm almost has no improvements compared with the Uniform baseline, which exactly demonstrates the sparsity of feedback. On average, the impression for each creative is less than one per day. The contextual algorithms have better performances where the interaction between different features are paid attention to capture the commonality between different creatives.

On the online A/B test, although the AutoCO is merely comparable in the first two days, the significant improvement is achieved on day 3. In the following consecutive three days, the AutoCO steadily improve the performance. This indicates the quick convergence of the proposed method which is essential for online advertisement. Compared with the FM-based models, our approach still shows competitive performances. It can be inferred that the search of interaction functions help find out more appropriate interaction functions to improve the expressiveness.

7 CONCLUSION

In this paper, we propose an automated creative optimization framework to model complex interaction between creative elements and to strike a balance between exploration and exploitation. For modeling complex interaction, we suggest an one-shot search algorithm for searching more effective interaction function in the customized

function space efficiently. For balancing between exploitation and exploration, we develop the stochastic variational inference for posterior approximation based on the reparameterization trick, and then apply Thompson Sampling for effective exploration of potentially better creatives. We conduct both offline and online experiments, showing that our proposed approach performs better than the competing baselines and verifying the effectiveness of the proposed algorithms.

ACKNOWLEDGMENTS

The work was supported by Alibaba Group through Alibaba Innovative Research Program. Defu Lian is supported by grants from the National Natural Science Foundation of China (Grant No. 61976198, 62022077), the National Key R&D Program of China under Grant No. 2020AAA0103800 and the Fundamental Research Funds for the Central Universities. Kai Zheng is supported by NSFC (No. 61972069, 61836007, 61832017) and Sichuan Science and Technology Program under Grant 2020JDTD0007.

REFERENCES

- [1] Shipra Agrawal and Navin Goyal. 2013. Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*. 127–135.
- [2] Javad Azimi, Ruofei Zhang, Yang Zhou, Vidhya Navalpakkam, Jianchang Mao, and Xiaoli Fern. 2012. The impact of visual appearance on user response in online display advertising. In *Proceedings of the 21st International Conference on World Wide Web*. 457–458.
- [3] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. 2015. Weight uncertainty in neural networks. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning-Volume 37*. 1613–1622.
- [4] Yin-Wen Chang, Cho-Jui Hsieh, Kai-Wei Chang, Michael Ringgaard, and Chih-Jen Lin. 2010. Training and testing low-degree polynomial data mappings via linear SVM. *Journal of Machine Learning Research* 11, 4 (2010).
- [5] Junxuan Chen, Baigui Sun, Hao Li, Hongtao Lu, and Xian-Sheng Hua. 2016. Deep ctr prediction in display advertising. In *Proceedings of the 24th ACM international conference on Multimedia*. 811–820.
- [6] Haibin Cheng, Roelof van Zwol, Javad Azimi, Eren Manavoglu, Ruofei Zhang, Yang Zhou, and Vidhya Navalpakkam. 2012. Multimedia features for click prediction of new ads in display advertising. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. 777–785.
- [7] Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. 2011. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. 208–214.
- [8] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: a factorization-machine based neural network for CTR prediction. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. 1725–1731.

- [9] Ruining He and Julian McAuley. 2016. VBPR: visual Bayesian Personalized Ranking from implicit feedback. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. 144–150.
- [10] Daniel N Hill, Houssam Nassif, Yi Liu, Anand Iyer, and SVN Vishwanathan. 2017. An efficient bandit algorithm for realtime multivariate optimization. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1813–1821.
- [11] Xian-Sheng Hua. 2018. Challenges and Practices of Large Scale Visual Intelligence in the Real-World. In *Proceedings of the 26th ACM international conference on Multimedia*. 364–364.
- [12] Frank Hutter, Lars Kotthoff, and Joaquin Vanschoren. 2019. *Automated machine learning: methods, systems, challenges*. Springer Nature.
- [13] Manas R Joglekar, Cong Li, Mei Chen, Taibai Xu, Xiaoming Wang, Jay K Adams, Pranav Khaitan, Jiahui Liu, and Quoc V Le. 2020. Neural input search for large scale recommendation models. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2387–2397.
- [14] Yuchin Juan, Yong Zhuang, Wei-Sheng Chin, and Chih-Jen Lin. 2016. Field-aware factorization machines for CTR prediction. In *Proceedings of the 10th ACM Conference on Recommender Systems*. 43–50.
- [15] Diederik P Kingma and Max Welling. 2014. Auto-Encoding Variational Bayes. *stat* 1050 (2014), 1.
- [16] Ron Kohavi. 1996. Scaling up the accuracy of Naive-Bayes classifiers: a decision-tree hybrid. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. 202–207.
- [17] Linghai Kong, Jimeng Sun, and Chao Zhang. 2020. SDE-Net: Equipping Deep Neural Networks with Uncertainty Estimates. *arXiv:cs.LG/2008.10546*
- [18] Jianan Li, Jimei Yang, Aaron Hertzmann, Jianming Zhang, and Tingfa Xu. 2018. LayoutGAN: Generating Graphic Layouts with Wireframe Discriminators. In *International Conference on Learning Representations*.
- [19] Lihong Li, Wei Chu, John Langford, Taesup Moon, and Xuanhui Wang. 2012. An unbiased offline evaluation of contextual bandit algorithms with generalized linear models. In *Proceedings of the Workshop on On-line Trading of Exploration and Exploitation* 2. 19–36.
- [20] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*. 661–670.
- [21] Defu Lian, Qi Liu, and Enhong Chen. 2020. Personalized ranking with importance sampling. In *Proceedings of The Web Conference 2020*. 1093–1103.
- [22] Defu Lian, Haoyu Wang, Zheng Liu, Jianxun Lian, Enhong Chen, and Xing Xie. 2020. Lightrec: A memory and search-efficient recommender system. In *Proceedings of The Web Conference 2020*. 695–705.
- [23] Defu Lian, Yongji Wu, Yong Ge, Xing Xie, and Enhong Chen. 2020. Geography-aware sequential location recommendation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2009–2019.
- [24] Bin Liu, Chenxu Zhu, Guilin Li, Weinan Zhang, Jincai Lai, Ruiming Tang, Xiquang He, Zhenguo Li, and Yong Yu. 2020. AutoFIS: Automatic Feature Interaction Selection in Factorization Models for Click-Through Rate Prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20)*. Association for Computing Machinery, New York, NY, USA, 2636–2645. <https://doi.org/10.1145/3394486.3403314>
- [25] Chenxi Liu, Barret Zoph, Maxim Neumann, Jonathon Shlens, Wei Hua, Li-Jia Li, Li Fei-Fei, Alan Yuille, Jonathan Huang, and Kevin Murphy. 2018. Progressive neural architecture search. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 19–34.
- [26] Hu Liu, Jing Lu, Hao Yang, Xiwei Zhao, Sulong Xu, Hao Peng, Zehua Zhang, Wenjie Niu, Xiaokun Zhu, Yongjun Bao, et al. 2020. Category-Specific CNN for Visual-aware CTR Prediction at JD. com. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2686–2696.
- [27] Hanxiao Liu, Karen Simonyan, Oriol Vinyals, Chrisantha Fernando, and Koray Kavukcuoglu. 2018. Hierarchical Representations for Efficient Architecture Search. In *International Conference on Learning Representations*.
- [28] Hanxiao Liu, Karen Simonyan, and Yiming Yang. 2018. DARTS: Differentiable Architecture Search. In *International Conference on Learning Representations*.
- [29] Kaixiang Mo, Bo Liu, Lei Xiao, Yong Li, and Jie Jiang. 2015. Image feature learning for cold start problem in display advertising. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.
- [30] Junwei Pan, Jian Xu, Alfonso Lobos Ruiz, Wenliang Zhao, Shengjun Pan, Yu Sun, and Quan Lu. 2018. Field-weighted factorization machines for click-through rate prediction in display advertising. In *Proceedings of the 2018 World Wide Web Conference*. 1349–1357.
- [31] Steffen Rendle. 2010. Factorization machines. In *2010 IEEE International Conference on Data Mining*. IEEE, 995–1000.
- [32] Carlos Riquelme, George Tucker, and Jasper Snoek. 2018. Deep Bayesian Bandits Showdown: An Empirical Comparison of Bayesian Deep Networks for Thompson Sampling. In *International Conference on Learning Representations*.
- [33] Jeff Schlummer. 1981. Mushroom records drawn from the audubon society field guide to north american mushrooms. *GH Lincoff (Pres), New York* (1981).
- [34] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.
- [35] William R Thompson. 1933. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* 25, 3/4 (1933), 285–294.
- [36] Sreekanth Vempati, Korah T Malayil, Sruthi V, and Sandeep R. 2019. Enabling Hyper-Personalisation: Automated Ad Creative Generation and Ranking for Fashion e-Commerce. *arXiv:cs.IR/1908.10139*
- [37] Yue Wang, Dawei Yin, Luo Jie, Pengyuan Wang, Makoto Yamada, Yi Chang, and Qiaozhu Mei. 2016. Beyond ranking: Optimizing whole-page presentation. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. 103–112.
- [38] Sirui Xie, Hehui Zheng, Chunxiao Liu, and Liang Lin. 2018. SNAS: stochastic neural architecture search. In *International Conference on Learning Representations*.
- [39] Xiao Yang, Tao Deng, Weihan Tan, Xutian Tao, Junwei Zhang, Shouke Qin, and Zongyao Ding. 2019. Learning Compositional, Visual and Relational Representations for CTR Prediction in Sponsored Search. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 2851–2859.
- [40] Quanming Yao, Xiangning Chen, James T Kwok, Yong Li, and Cho-Jui Hsieh. 2020. Efficient neural interaction function search for collaborative filtering. In *Proceedings of The Web Conference 2020*. 1660–1670.
- [41] Quanming Yao, Ju Xu, Wei-Wei Tu, and Zhanxing Zhu. 2020. Efficient Neural Architecture Search via Proximal Iterations. In *AAAI*. 6664–6671.
- [42] Zhichen Zhao, Lei Li, Bowen Zhang, Meng Wang, Yuning Jiang, Li Xu, Fengkun Wang, and Weiyang Ma. 2019. What You Look Matters? Offline Evaluation of Advertising Creatives for Cold-start Problem. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 2605–2613.
- [43] Zhao Zhong, Junjie Yan, Wei Wu, Jing Shao, and Cheng-Lin Liu. 2018. Practical block-wise neural network architecture generation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2423–2432.
- [44] 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.
- [45] Barret Zoph and Quoc Le. 2016. Neural Architecture Search with Reinforcement Learning. (2016).
- [46] Barret Zoph and Quoc V Le. 2017. Neural Architecture Search with Reinforcement Learning. *arXiv:cs.LG/1611.01578*
- [47] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V Le. 2018. Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 8697–8710.