

The Whole-Page Optimization via Dynamic Ad Allocation

Weiru Zhang, Chao Wei, Xiaonan Meng, Yi Hu, Hao Wang

Alibaba Group
Hangzhou, China

{weiru.zwr, weichao.wc, xiaonan.mengxn, erwin.huy, longran.wh}@alibaba-inc.com

ABSTRACT

Modern search engines present result pages composed of two most prominent types of information: sponsored and organic search results. The whole-page results must satisfy user's information inquiry while sponsored ad alongside the search results has become a key monetization strategy for the platform. Against the backdrop of this situation, a basic question has received comparatively little attention: how many ads are good enough to get higher user satisfaction and better monetization? Most search engines always display a fixed number of ads or use heuristic rules to determine the number of ads.

In this paper, we formulate the task of finding the best number of ads into a linear programming optimization problem, for which we propose a novel online algorithm to solve. We have conducted several offline experiments and tested our approach in Alibaba E-commerce platform. The experimental results show that the platform could achieve higher revenue and more clicks simultaneously by the proposed algorithm.

CCS CONCEPTS

• **Information systems** → **Information retrieval**; • **Applied computing** → **Electronic commerce**; • **Computing methodologies** → **Machine learning**;

KEYWORDS

sponsored search, whole-page optimization, dynamic ad allocation, click yield, deep neural network

ACM Reference Format:

Weiru Zhang, Chao Wei, Xiaonan Meng, Yi Hu, Hao Wang. 2018. The Whole-Page Optimization via Dynamic Ad Allocation. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3184558.3191584>

1 INTRODUCTION

When a user searches a query, the search engine returns a page consisting of several ads followed by search results. Both ads and organic search results are necessary and provide information to users. Based on different purposes, the two types of results are recalled and optimized separately. Organic results are often ranked by the factors like relevance or the popularity of the item for the query. Advertisers can get the top rank in sponsored search by giving high bids, so ad results are determined based on the potential revenue

for the search engine. Under the pay per click (PPC) advertising model, the total revenue is the number of clicks multiplied by the cost per click (CPC). How many ads are enough? One intuition is that more ads could lead higher revenue for the search engine, but may impact user experience. One fundamental challenge facing all search engines is how to achieve a balance between revenue and user experience. Most search engines still display a fixed number of ads or use heuristic rules to determine the number of ads.

This paper proposes a solution of dynamic ad allocation to optimize the whole-page performance. Specifically, we define a new optimization problem which involves the whole-page items, considering the trade-off between revenue and relevance of the whole-page as indicated by the total number of clicks. Furthermore, A model based on deep neural network (DNN) is proposed to predict the ratio of the total number of clicks over the total number of impressions (*click yield, CY*) with taking into account the interaction between each items.

2 RELATED WORK

Previous work largely focused on studying optimization of sponsored and search results separately. There have been many work related to revenue optimization [2, 10] and search experience optimization [3]. In addition, some research has been so far limited to study the interplay between the two types of search results. Typically, the authors in [1, 13, 14, 20] find that clicks on organic search results have a complementary effect on the clicks on sponsored ads, while in [8, 19], the authors investigate the relationship between the perceived usefulness of the ads and their textual similarity to the organic results based on the classification of queries. The conclusion in [8] demonstrates that for navigational queries there is a clear competition between ads and organic results, while for non-navigational queries the competition turns into synergy. A general presentation optimization framework is proposed to evaluate user's satisfaction in human computer interaction area [17, 26]. These findings allow us to better understand the relationship between sponsored and organic results.

We formulate the whole-page optimization as a new allocation problem, which extends the online ad allocation problem proposed by [23]. Traditional ad allocation problem includes the Display Ad Allocation problem [2, 5, 9, 12, 28] and AdWords problem [6, 18, 21, 29]. In addition, an online algorithm for the video-ad allocation problem [24] extends the ad-auction allocation framework [15]. In these problems, online request must be assigned to an inventory of ads, optimizing the total revenue of the allocations while respecting pre-specified contracts.

Some of the other related work [29] proposes a trade-off ranking model to deliver high quality ads to the user as well as maximize search engine revenue. Ad number prediction is discussed only tentatively in advertising scope [16, 25].

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '18 Companion, April 23–27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

<https://doi.org/10.1145/3184558.3191584>

3 PRELIMINARIES

This paper studies the problem of optimizing the number of sponsored items to be displayed along organic items. There needs to be a trade-off between revenue and relevance of the whole-page items as indicated by the total number of clicks. We formulate the problem as an allocation problem which maximizes the expected revenue with *click yield* constraint.

3.1 Click Yield

Compared to individual click-through rate (*ctr*), we use the concept of *click yield* proposed by [27]. Given a specific request i and a list j which represents the page allocation, j contains N items including ads and organic results $j = \{j_1, j_2, \dots, j_N\}$. The *click yield* is the ratio of the total number of clicks on all the items over the total number of impressions:

$$CY_{ij} = \frac{\sum_{n=1}^N Click(i, j_n)}{Imp(i, j)} \quad (1)$$

Click yield measures the overall performance of the list which must consider some challenges including ad depth effects, interactive influence rather than single ad or organic result.

3.2 Problem Definition

With the definition of CY , we formalize the whole-page optimization problem as follows:

$$\begin{aligned} & \max \sum_{i,j} REV_{ij} \cdot x_{ij} \\ & s.t. \frac{\sum_{i,j} CY_{ij} \cdot x_{ij}}{\sum_{i,j} x_{ij}} \geq T, \\ & \quad \forall i, \sum_j x_{ij} = 1, \\ & \quad \forall i, j, x_{ij} \geq 0. \end{aligned} \quad (2)$$

For each request i and a ranking list j , we let REV_{ij} and CY_{ij} denote the expected revenue and overall click performance of the page while j is assigned as a return list to request i . This problem definition aims to find the allocation plan of x_{ij} which maximizes the revenue under the average CY constraint T . x_{ij} is a binary variable which equals to one if j is assigned to request i , otherwise zero. Only one result list will be chosen for each request, so we set $\sum_j x_{ij} = 1$.

The dual of (2) can be written as follows:

$$\begin{aligned} & \min \sum_i \beta_i \\ & s.t. \forall i, j, \beta_i \geq REV_{ij} + \alpha \cdot (CY_{ij} - T) \\ & \quad \alpha \geq 0 \end{aligned} \quad (3)$$

The parameter α arises naturally as the dual variable of the constraint enforcing a lower bound on the average CY in the global offline linear programming. It can be interpreted as the trade-off between revenue and click-through rate of the whole-page. Moreover, the dual problem can be solved by the linear programming package and α will be used in the online ranking algorithm in Section 4.1. β is the shadow revenue.

4 METHODOLOGY

4.1 The Real-Time Algorithm

Inspired by the work of Chen *et al.* [5], we develop a new real-time algorithm to solve the revenue maximization problem with constraints via varying the number of ads dynamically for each request. The online algorithm is referred to as K-Algorithm and shown in Algorithm 1.

Algorithm 1 Real-Time K-Algorithm

Input: α, N
Output: x_{ij}, k

```

1: for request  $i$  from an online stream do
2:    $J \leftarrow \emptyset$ 
3:    $A \leftarrow AdRetrieveRank(i)$ 
4:    $O \leftarrow OrganicRetrieveRank(i)$ 
5:   for  $k = 0$  to  $N$  do
6:      $J \leftarrow J \cup [A(1), A(2), \dots, A(k), O(1), \dots, O(N-k)]$ 
7:   end for
8:   for  $j \in J$  do
9:      $CY_{ij} = predictCY(i, j)$ 
10:     $REV_{ij} = predictREV(i, j)$ 
11:   end for
12:    $j' \leftarrow \operatorname{argmax}_{j \in J} (REV_{ij} + \alpha \cdot CY_{ij})$ 
13:    $x_{ij'} = 1$ 
14:    $x_{ij} = 0, \forall j \neq j'$ 
15: end for
```

For an incoming request i , candidate sets of sponsored ads and organic items are retrieved and ranked separately. Then a return list is generated by concatenating top k ads with top $N-k$ organic items (line 6). k is the number of ads which determines the final ranking list j for each specific request i . In this case when $k=0$, j is an item list without ad exposure. It's worth noting that the combination form is not always fixed as the top- k ads layout and depends on the page layout. Then we calculate CY_{ij} and REV_{ij} for each impression (i, j) which will be introduced in Section 4.2. We compare each ranking list j for request i with the value of $REV_{ij} + \alpha \cdot CY_{ij}$ (a simplified form of $REV_{ij} + \alpha \cdot (CY_{ij} - T)$), and the j with the maximal value will be chosen (line 12-14). The input α can be calculated in advance with the offline search log data by Equation (3).

4.2 Predicting CY and REV

Obviously CY and REV play key roles in the algorithm. To insure the efficiency of the algorithm, we try to estimate them as accurately as possible. In this subsection, we discuss two different estimation methods.

4.2.1 Sum Method. Current ad-delivery strategy often predicts individual *ctr* and ignores correlation between different items. An intuitive solution is to directly add *ctr* of each item as follows:

$$CY_{ij} = \sum_{n=1}^N ctr_{ijn} \quad (4)$$

where ctr_{ij_n} is the estimated click-through rate of the n -th item in list j for request i . Furthermore, in consideration of position bias, we estimate CY with *Examination Hypothesis* [4, 7, 11]. *Examination Hypothesis* says that users are more likely to click the first rank item and less likely to look at items in lower ranks, which suggests that each rank has a certain probability of being examined. This probability is denoted by $P(e|n)$, ctr_{ij_n} is written:

$$ctr_{ij_n} = P(e|n) \cdot P(c|i, j_n) \quad (5)$$

$P(c|i, j_n)$ is the click-through rate of j_n which is an element of result j for request i . Similarly, we have REV_{ij} :

$$REV_{ij} = \sum_{n=1}^N rpm_{ij_n} = \sum_{n=1}^N ctr_{ij_n} \cdot clkprice_{ij_n} \quad (6)$$

REV_{ij} depends on the auction model. Under the pay per click advertising model, the total revenue is the number of clicks multiplied by the cost per click. In addition, the search engines commit to using generalized second price (GSP) auction mechanism [22] for determining how much each advertiser should pay. When the list j is fixed, we can easily use expected click-through rate and real click price to calculate the expected revenue for every impression. For the organic items, $clkprice$ is set to zero.

4.2.2 Interactive Method. Estimating CY is a more complicated problem than estimating individual item ctr . *Sum Method* is easy to implement with position bias, but it doesn't take into account extra factors: interactive influences across items, especially the interaction between sponsored ads and organic items. Two systems that have similar precisions on predicting individual performance could be very different on group performance[27]. Thus we are going to seek a model to capture the interaction.

Here we want to use *DNN* to predict CY and REV . *DNN* is a feed-forward neural network and each hidden layer of the network is defined as:

$$l_{i+1} = f(W_i l_i + b_i) \quad (7)$$

where i is the hidden layer number and f is the activation function (like *ReLU*, *Sigmoid*). l_i , b_i and W_i are activation value vector, bias vector and model weights at i -th layer. The input features consist of multiple kinds of information such as performance (estimated ctr , estimated revenue per impression (rpm), item information (price, category, type (ad or organic item), etc.) and contextual information (the nearby items' information). The detail of feature and parameter setting will be introduced in 5.2.

5 EXPERIMENTS

In this section, we conduct comprehensive experiments to validate the proposed algorithm. The offline simulation is based on a click-through log data from alibaba.com, which is currently the world's largest cross border business to business e-commerce website. We also run online A/B tests on Alibaba E-commerce platform using the *Sum Method* described in 4.2.1. In the last part of Section 5.2, compared to *Sum Method*, *Interactive Method* shows higher accuracy in *click yield* prediction.

5.1 Offline Simulation

We build an offline simulation platform and replay the request stream. We use one weekly log data including sponsored ads and organic items collected from alibaba mobile in September, 2017. The data after filtering covers 7.7 million requests. Each request corresponds to a candidate set that contains at most ten ads and twenty organic items. The baseline is a fixed setting of up to five ads on the top of the page and the rest are organic items which keeps consistent with the online setting. Note that since there may not be enough candidate ads for some specified requests, the average show number (ASN) of ads may be less than five.

5.1.1 Solution Space of Parameter T . Our empirical study focuses on controlling two core Key Performance Indicators (KPIs): CY and REV . The Linear programming problem (2) defined in Section 3.2 shows that the constraint T (the average CY) determines the allocation plan (K ads for each request) and the overall maximal revenue, which suggests that the parameter T is capable of settling the revenue KPI at the given reference value. In other words, we can leverage the T controller as a revenue optimization tool and study the performance on revenue optimization by searching the parameter space of T . The revenue optimization problem is a monotone decreasing function of T on account of the property of linear programming. Furthermore, considering that what we want is to make the two KPI s better (or at least not worse), the parameter T definitely has the upper and lower bounds. Obviously, the lower bound T_{lower} is the average CY of baseline. And T_{upper} stands for the upper bound of T which yields the same revenue as the baseline. Figure 1 shows the controlled revenue curve against the parameter T . The red dot in Figure 1 is the baseline with a fixed five ads setting for each request. From the result, we can see that the interval (T_{lower}, T_{upper}) is our solution space for the two core KPI s.

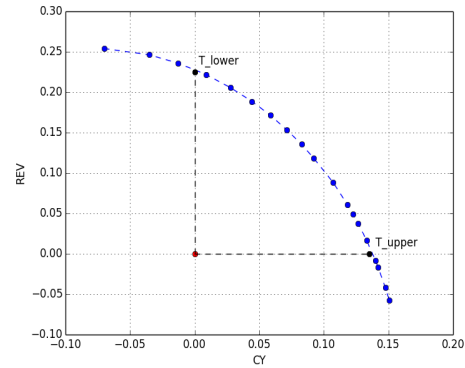


Figure 1: Overall revenue improvement under different cy constraint. All measurements are calculated relative to the baseline.

5.1.2 Parameter T Tuning. In this subsection, we share offline simulation results about parameter T tuning in the solution space. In addition to CY and REV , we study several other KPI s: average

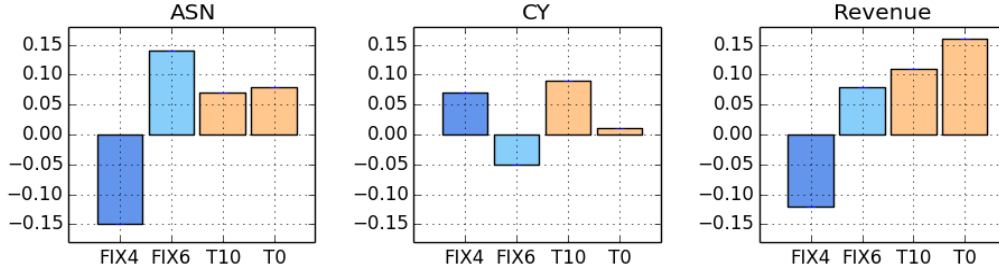


Figure 2: Relative performance for online evaluation

show number of sponsored ads (ASN), the average item ctr of ads (ctr_{ad}), the average item ctr of organic results (ctr_{org}).

Since the train/test data are split temporally rather than randomly, the problem of predicting the past using future data is avoided. We adapt the controllers over the test data and observe the performance. The detailed control performance on each parameter T is provided in Table 1 for all *KPIs*. We also compare our algorithm with the fixed number strategy (*Fixed*). We run experiments with four different parameter T settings in K-algorithm. The number after T stands for the percentage of improvement on average CY , e.g. $T10$ represents the value of parameter T is 10% improvement of the baseline.

Table 1: Relative performance for offline simulation

	ASN	REV	CY	ctr_{ad}	ctr_{org}
Fixed4	-16.3%	-13.8%	10.2%	10.9%	9.4%
Fixed6	18.1%	8.4%	-7.0%	-7.2%	-6.8%
T-5	12.5%	25.1%	-5.0%	-6.3%	-3.8%
T0	9.2%	22.5%	0.0%	4.7%	-4.2%
T5	7.9%	17.4%	5.0%	7.0%	3.1%
T10	7.2%	10.1%	10.0%	12.3%	8.1%

From the results we can see the algorithm proposed outperforms the fixed strategies, which verifies the effectiveness of K-algorithm with different parameter T settings. The results are consistent with the analysis of T solution space in Section 5.1.1.

5.2 Online Evaluation

We implement our algorithm in alibaba.com and run experiments on live traffic. We test on two different channels in mobile during six consecutive days in October 2017. The baseline is *Fixed5* which is the online setting. The performance comparison with various measures are reported in Figure 2. Due to the data sensitivity, we only present the relative performance here.

From the comparison we have the following observations: (1) Both *REV* and *CY* of the proposed algorithm are better than the baseline. *T10* outperforms *T0* in *CY* while *T0* gets better *REV* than *T10*. The result is consistent with the analysis in Section 5.1. (2) Compared with *Fixed6*, the proposed algorithm gets both higher *REV* and *CY*. It means dynamic ad allocation is very powerful, we can get significant improvement in both *CY* and *REV* without changing the ranking strategies of ads and organic items or largely

increasing the total number of ads. (3) We also notice that the results of live traffic are significantly less than the results in the simulations. Since we use estimated *CY* and *REV* without taking into account the interaction, there is still a gap between target $CY(T)$ and real *CY*.

We have implemented a simple *DNN* model with two hidden layers and *ReLU* as the activation function by *Interactive Method* described in 4.2.2 to calculate *CY*. The features are only individual *ctrs* of each item. We compare the *DNN* model with *Sum Method* with and without considering position bias. Individual *ctrs* are generated by a logistic regression model. The comparison results on MSE are shown in Table 2. We would like to point out that even the network structure and types of feature are very simple, the predictive result of *Interactive Method* is still better than *SUM Method*. In the future work, more complex features and models will be included.

Table 2: Predictive results of *CY* prediction

	<i>SUM</i>	<i>SUM_{pos_bias}</i>	<i>DNN</i>
MSE	0.1169	0.1129	0.1116

6 CONCLUSION AND FUTURE WORK

In this paper, we study the problem of whole-page optimization via dynamic ad allocation. Result pages of major search engines include information from sponsored ads and organic search results. The whole-page results must satisfy user's information inquiry while sponsored ads alongside the search results profit the search engine. To the best of our knowledge, previous work focused on studying optimization of the specified targets for ads and organic items separately.

We address a new optimization problem maximizing search engine revenue as well as improving user experience, which extends the general allocation framework. Experiments on the real data including offline simulation and online evaluation in Alibaba E-commerce platform demonstrate that the algorithm we proposed gets significant improvements in both revenue and click-through rate.

In the future work, we will study *click yield* and *revenue* measurements and focus on predicting group performance for a ranking list instead of individual performance. Several factors like ad depth, interaction between different elements in the whole-page will be considered.

REFERENCES

- [1] Ashish Agarwal, Kartik Hosanagar, and Michael Smith. 2011. Sponsored Search: Do Organic Results help or hurt the Performance and under what conditions? (2011).
- [2] Anand Bhalgat, Jon Feldman, and Vahab Mirrokni. 2012. Online allocation of display ads with smooth delivery. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1213–1221.
- [3] Fidel Cacheda, Nicola Barbieri, and Roi Blanco. 2017. Click Through Rate Prediction for Local Search Results. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 171–180.
- [4] Olivier Chapelle and Ya Zhang. 2009. A dynamic bayesian network click model for web search ranking. In *Proceedings of the 18th international conference on World wide web*. ACM, 1–10.
- [5] Ye Chen, Pavel Berkhin, Bo Anderson, and Nikhil R Devanur. 2011. Real-time bidding algorithms for performance-based display ad allocation. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1307–1315.
- [6] Alexey Chervonenkis, Anna Sorokina, and Valery A Topinsky. 2013. Optimization of ads allocation in sponsored search. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 121–122.
- [7] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008. An experimental comparison of click position-bias models. In *Proceedings of the 2008 international conference on web search and data mining*. ACM, 87–94.
- [8] Cristian Danescu-Niculescu-Mizil, Andrei Z Broder, Evgeniy Gabrilovich, Vanja Josifovski, and Bo Pang. 2010. Competing for users' attention: on the interplay between organic and sponsored search results. In *Proceedings of the 19th international conference on World wide web*. ACM, 291–300.
- [9] Nikhil R Devanur, Zhiyi Huang, Nitish Korula, Vahab S Mirrokni, and Qiqi Yan. 2016. Whole-page optimization and submodular welfare maximization with online bidders. *ACM Transactions on Economics and Computation* 4, 3 (2016), 14.
- [10] Nikhil R Devanur, Kamal Jain, Balasubramanian Sivan, and Christopher A Wilkens. 2011. Near optimal online algorithms and fast approximation algorithms for resource allocation problems. In *Proceedings of the 12th ACM conference on Electronic commerce*. ACM, 29–38.
- [11] Georges E Dupret and Benjamin Piwowarski. 2008. A user browsing model to predict search engine click data from past observations.. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 331–338.
- [12] Jon Feldman, Monika Henzinger, Nitish Korula, Vahab S Mirrokni, and Cliff Stein. 2010. Online stochastic packing applied to display ad allocation. In *European Symposium on Algorithms*. Springer, 182–194.
- [13] Anindya Ghose and Sha Yang. 2008. Comparing performance metrics in organic search with sponsored search advertising. In *Proceedings of the 2nd International Workshop on Data Mining and Audience Intelligence for Advertising*. ACM, 18–26.
- [14] Anindya Ghose and Sha Yang. 2009. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science* 55, 10 (2009), 1605–1622.
- [15] Ashish Goel, Mohammad Mahdian, Hamid Nazerzadeh, and Amin Saberi. 2010. Advertisement allocation for generalized second-pricing schemes. *Operations Research Letters* 38, 6 (2010), 571–576.
- [16] Sreenivas Gollapudi, Rina Panigrahy, and Moises Goldszmidt. 2011. Inferring clickthrough rates on ads from click behavior on search results. In *Proceedings of the Workshop on User Modeling for Web Applications, Fourth International Conference on Web Search and Web Data Mining, WSDM*.
- [17] Laura A Granka, Thorsten Joachims, and Geri Gay. 2004. Eye-tracking analysis of user behavior in WWW search. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 478–479.
- [18] Di He, Wei Chen, Liwei Wang, and Tie-Yan Liu. 2013. A Game-Theoretic Machine Learning Approach for Revenue Maximization in Sponsored Search.. In *IJCAI* 206–212.
- [19] Bernard Jansen and Amanda Spink. 2007. The effect on click-through of combining sponsored and non-sponsored search engine results in a single listing. In *Proceedings of the 2007 Workshop on Sponsored Search Auctions, WWW Conference*.
- [20] Bernard J Jansen and Marc Resnick. 2005. Examining searcher perceptions of and interactions with sponsored results. In *Workshop on Sponsored Search Auctions*.
- [21] Chinmay Karande, Aranyak Mehta, and Ramakrishnan Srikant. 2013. Optimizing budget constrained spend in search advertising. In *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 697–706.
- [22] Mervyn King, Jill Atkins, and Michael Schwarz. 2007. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *The American economic review* 97, 1 (2007), 242–259.
- [23] Aranyak Mehta et al. 2013. Online matching and ad allocation. *Foundations and Trends® in Theoretical Computer Science* 8, 4 (2013), 265–368.
- [24] Hanna Sumita, Yasushi Kawase, Sumio Fujita, Takuro Fukunaga, and RIKEN AIP Center. Online Optimization of Video-Ad Allocation. (????).
- [25] Bo Wang, Zhaonan Li, Jie Tang, Kuo Zhang, Songcan Chen, and Liyun Ru. 2011. Learning to advertise: how many ads are enough? *Advances in Knowledge Discovery and Data Mining* (2011), 506–518.
- [26] 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*. ACM, 103–112.
- [27] Dawei Yin, Bin Cao, Jian-Tao Sun, and Brian D Davison. 2014. Estimating ad group performance in sponsored search. In *Proceedings of the 7th ACM international conference on Web search and data mining*. ACM, 143–152.
- [28] Weinan Zhang, Shuai Yuan, and Jun Wang. 2014. Optimal real-time bidding for display advertising. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1077–1086.
- [29] Yunzhang Zhu, Gang Wang, Junli Yang, Dakan Wang, Jun Yan, and Zheng Chen. 2009. Revenue optimization with relevance constraint in sponsored search. In *Proceedings of the Third International Workshop on Data Mining and Audience Intelligence for Advertising*. ACM, 55–60.