



**HAL**  
open science

# Planning-based Approach for Optimizing the Display of Online Advertising Campaigns

Sertan Girgin, Jérémie Mary, Philippe Preux, Olivier Nicol

► **To cite this version:**

Sertan Girgin, Jérémie Mary, Philippe Preux, Olivier Nicol. Planning-based Approach for Optimizing the Display of Online Advertising Campaigns. NIPS workshop on Machine Learning in Online Advertising, Dec 2010, Whistler, Canada. hal-00772512

**HAL Id: hal-00772512**

**<https://hal.science/hal-00772512>**

Submitted on 10 Jan 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# Planning-based Approach for Optimizing the Display of Online Advertising Campaigns

---

Sertan Girgin, Jeremie Mary, Philippe Preux and Olivier Nicol  
Team-Project SequeL  
INRIA Lille Nord Europe, and LIFL (UMR CNRS), Université de Lille, France  
name.surname@inria.fr

## Abstract

In a realistic context, the online advertisements have constraints such as a certain number of clicks to draw, as well as a lifetime. Furthermore, receiving a click is usually a very rare event. Thus, the problem of choosing which advertisement to display on a web page is inherently dynamic, and intimately combines combinatorial and statistical issues. We introduce a planning based algorithm for optimizing the display of advertisements and investigate its performance through simulations on a realistic model designed with an important commercial web actor.

## 1 Introduction and formalization of the problem

In this paper, we consider the problem of selecting advertisements in order to maximize the revenue earned from clicks in the “cost per click” economic model under different settings. Our goal is not to optimize any asymptotic behavior or exhibit algorithms that are able to achieve optimal asymptotic behavior, but rather to solve efficiently the practical problem that arises on a web site and involves certain degrees of uncertainty originating from various sources.

We define the problem as follows. At a given time  $t$ , there is a pool of  $K$  advertising campaigns denoted by  $K^t$ . Each campaign  $Ad_k \in K^t$  is characterized by a tuple  $(s_k, S_k, L_k, B_k, b_k^t, r_k)$  where  $k$  is the identifier of the campaign,  $s_k, S_k, L_k, B_k$  and  $r_k$  are its status, starting time, lifetime and total click budget and the revenue obtained per click respectively.  $b_k^t \leq B_k$  denotes the remaining budget of the campaign at time  $t$ . The campaign lasts for  $L_k$  time steps and expects to receive  $B_k$  clicks during its lifetime. The status of an advertising campaign can be either: **scheduled** when the campaign will begin at some time in the future, **running** when the campaign is active (*i.e.*  $S_k \leq t < S_k + L_k$ ), or **expired** when the campaign has ended (*i.e.*  $S_k + L_k \leq t$  or  $b_k^t = 0$ ). Only the advertisements of running campaigns can be displayed. The web server receives a continuous stream of visitors, each of which is assumed to be from one of  $N$  possible user profiles. The probability that the visitor belongs to a certain profile  $P_i$  is  $R_i$  with  $\sum_{i=1}^N R_i = 1$ . When a visitor visits the web site, a new “session” begins and we observe one or several iterations of the following sequence of events: (i) the visitor requests a certain page at time  $t$  (ii) the requested page is displayed to this visitor with an advertisement  $Ad_k$  embedded in it, (iii) the visitor clicks on the advertisement with probability  $p_{i,k}$  where  $i$  denotes the user profile of the visitor; this probability is usually called the *click-through rate* (CTR), (iv) if there is a click, then the revenue associated with the advertisement  $r_k$  is incurred. After a certain number of page requests, the visitor leaves the web site and the session terminates. The objective is to maximize the total revenue by choosing the advertisements to be displayed “carefully”. Since page requests are atomic actions, in the rest of the paper we will take a page request as the *unit of time* to simplify the discussion.

In the simplest case, we assume that (a) time horizon  $T$  is fixed, (b) the pool of advertising campaigns at each time step is given, (c) the visit probabilities of user profiles,  $R_i$ , and their click probabilities for each campaign,  $p_{i,k}$ , and the profile of each visitor are known. Note that, the visitor at time  $t$  and

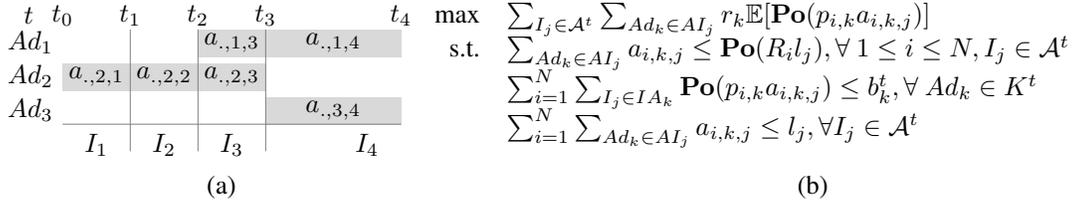


Figure 1: (a) The timeline divided into intervals and parts.  $Ad_{1,3}$  are in scheduled state at time  $t_1$ , and  $Ad_2$  expire after  $t_3$ .  $I_j$  denotes the  $j^{\text{th}}$  interval  $[t_{j-1}, t_j]$  and  $a_{i,k,j}$  denotes the allocation for  $Ad_k$  for users belonging to profile  $R_i$  in interval  $I_j$ . The first index is dropped for the sake of clarity. (b) Stochastic formulation of the linear program.  $AI_j$  denotes the set of running campaigns in interval  $I_j$ ,  $l_j$  is the length of interval  $I_j$ , and  $IA_k$  denotes the set of intervals that cover  $Ad_k$ .

whether he will click on the displayed advertisement or not are still unknown. Under this setting, given a visitor from profile  $P_i$  at time  $t$ , one possible and efficient way to choose an advertising campaign is to use the *highest expected value* (HEV) policy and pick the running campaign with the highest expected revenue per click, *i.e.*  $\text{argmax}_{Ad_k \in K^t} r_k p_{i,k}$ . Alternatively, one can employ the *stochastic expected value* (SEV) policy in which the selection probability of a running campaign is proportional to its expected revenue per click. As both policies exploit advertising campaigns with possibly high return and assign lower priority to those with lower return, they are expected to perform well if the lifetimes of the campaigns are “long enough” to ensure their total click budgets. However, even under some simple cases they may perform inferior to choosing an advertisement randomly at each step (see the example in [3] Sec. 2.1.1). In order to do better, it is compulsory to take into consideration the interactions between the advertising campaigns which materialize as *overlapping time intervals* over the timeline (Fig. 1 (a)); the problem then becomes finding an *allocation* of the number of advertising campaign displays in each interval such that (a) the allocation for a particular user profile is not over the capacity of the interval, (b) the remaining total click budgets are not exceeded, and (c) the total expected revenue is maximized. This corresponds to the maximization of a *linear objective function* subject to *linear inequality constraints*, which is a *linear programming* problem and can be solved efficiently; the detailed formulation and discussion can be found in [3]. The solution of the linear program at time  $t$  indicates the number of displays that should be allocated to each campaign for each user profile and in each interval, but it does not provide a specific way to choose the campaign to display to a user from a particular profile at time  $t$ . For this purpose, it is possible use the ratio of displays allocated to a particular campaign to the total allocation of advertising campaign displays for that user profile in the corresponding interval. One can either pick the campaign having the highest ratio, called the *highest LP policy* (HLP), or employ the *stochastic LP policy* (SLP) in which the selection probability of a campaign is proportional to its ratio. The linear program can either be solved at each time step or if this option is not feasible (*e.g.* due to computation time constraints) with regular periods or intermittently (*e.g.* when the budget of a campaign is met). In the latter case, the resulting allocation is used to determine the campaigns to be displayed until the next problem instance is solved by updating the allocated number of campaign displays as we move along the timeline and reducing the allocation of the chosen campaigns in the corresponding intervals. The complete algorithm is presented in [3] Fig. 4.

The static setting with full information has two sources of uncertainty: (a) the user profiles of visitors are drawn from a categorical distribution, and (b) each campaign display is a Bernoulli trial with a certain probability, which is known, and the result is either a success (*i.e.* click) or a failure. The aforementioned linear programming solution of the optimization problem focuses on what happens in the expectation. Following the resulting policy in different realizations of the random problem may lead to different total revenue that vary from its expected value (see the example in [3] Sec. 2.1.2). In reality, reducing this variability may also be important and could be considered as another objective. Note that, the expected number of visitors from user profile  $P_i$  during the timespan of interval  $I_j$  and the expected number of clicks that would be received if the campaign  $Ad_k$  is displayed  $a_{i,k,j}$  times to the visitors from user profile  $P_i$  can be considered as random variables having Poisson distributions with parameters  $R_i t$  and  $p_{i,k} t$ , respectively. Let  $\mathbf{Po}(\lambda)$  denote a Poisson-distributed random variable with parameter  $\lambda$ . Replacing the corresponding terms in the linear program by the random variables, we obtain the stochastic optimization problem presented in Fig. 1 (b). The sum of

independent Poisson-distributed random variables also follows a Poisson distribution with parameter equal to the sum of their parameters. Assuming that  $\mathbf{Po}(p_{i,k}a_{i,k,j})$  are independent, the budget constraints can be written as  $\mathbf{Po}(\sum_{i=1}^N \sum_{I_j \in I_{A_k}} p_{i,k}a_{i,k,j}) \leq b_k^t, \forall Ad_k \in K^t$  which is equivalent to its linear program counterpart in expectation. The rationale behind this set of constraints is to bound the total expected number of clicks for each campaign, while at the same time trying to stay as close as possible to the bounds due to maximization in the objective function. Assume that in the optimal allocation the budget constraint of campaign  $Ad_k$  is met. This means that the expected total number of clicks for  $Ad_k$  will be a Poisson-distributed random variable with parameter  $b_k^t$  and in any particular instance of the problem the probability of realizing this expectation would be 0.5. In order to increase the likelihood of reaching the target expected total number of clicks, a possible option would be to use a higher budget limit. Let  $\alpha_k$  be our risk factor and  $\mathbf{Po}(\lambda_k)$  be the Poisson-distributed random variable having the smallest parameter  $\lambda_k$  such that  $Pr(\mathbf{Po}(\lambda_k) > b_k^t - 1) \geq \alpha_k$ ;  $b_k^t$  and  $\alpha_k$  are known, and  $\lambda_k$  can be found using numerical methods. If we replace  $b_k^t$  with  $\lambda_k$  in the budget constraint and solve the linear optimization problem again, the expected total number of clicks for  $Ad_k$  based on the new allocation would be greater than or equal to  $b_k^t$  and will have an upper bound of  $\lambda_k$ . Following the same strategy, one can derive new bounds for the user profile constraints and replace  $R_i l_j$  terms with the corresponding parameters of the random variables.

So far, we have assumed that the visit probabilities of user profiles and their click probabilities for each campaign are known. In reality, these probabilities are hardly known in advance and have to be estimated. By noting that we can consider them as categorical and Bernoulli random variables, respectively, it is possible to estimate their value by using maximum likelihood or Bayesian maximum a posteriori estimation with conjugate priors of Beta and Dirichlet distributions (see [3] Sec. 2.1.3). As we will see in the next section, in the latter case choosing good priors may have a significant effect on the outcome. By estimating probabilities at each step (or periodically) and replacing the actual values with their estimates, we can determine allocations (optimal up to the accuracy of the estimations) and choose advertising campaigns to display. For maximum a posteriori estimates, the mode of the posterior distribution can be used as a point estimate and a single instance of the problem can be solved, or several instances of the problem can be generated by sampling probabilities from the posterior distributions, solved separately and then the resulting allocations can be merged (e.g. by taking their mean; in this case the final allocations will likely be not bound to the initial constraints). As in many online learning problems, one important issue that arises here is the need for balancing the exploitation of the current estimates and exploration, i.e. focusing on less-certain (e.g., with higher variance) parameters; possible approaches are discussed in [3] Sec. 2.1.3.

In the more realistic dynamic setting, the time horizon is no longer fixed, and furthermore new campaigns may appear with time. We will consider two main cases in which either we have a *generative model* or not, which given a set of parameters and the current state can generate advertising campaigns during a specified time period. When a model is not available, only campaigns that have been revealed are known and they impose a certain maximum time horizon  $H_{max}$ . Although, it is possible to apply the proposed method and calculate the allocations for them, doing so would ignore the possibility of the arrival of new campaigns that may overlap and intervene with the existing ones; the resulting *long-term* policies may perform well if the degree of dynamism in the environment is not high. On the contrary, one can focus only on short or medium-term conditions omitting the known campaigns that start after a not-too-distant time  $H$  in the future. The resulting policies will be greedier as  $H$  is smaller and disregard the long-time interactions between the existing campaigns; however, they will also be less likely to be affected by the arrival of new campaigns (see the example in [3] Sec. 2.2). For such policies, choosing the optimal value of the planning horizon is not trivial due to the fact that it strongly depends on the underlying model. One possible way to remedy this situation would be to solve for a set of different planning horizons  $H_1, \dots, H_u = H_{max}$  (as the planning horizons differ, the structure of the optimization problems would also be different from each others) and then combine the resulting probability distributions of campaign displays, such as by majority voting. When a model is available, it can be utilized to compensate for the uncertainty in future events by allowing us to generate a set of *hypothetical* campaigns (for example, up to  $H_{max}$ ), simulating what may happen in future, and include them in the planning phase. By omitting allocations made for these hypothetical campaigns from the allocation scheme found by solving the optimization problem, display probabilities that inherently take into consideration the effects of future events can be calculated. Note that, this would introduce bias to the resulting policies which can be reduced by running multiple simulations and combining their results as mentioned before.

## 2 Experiments

Our approach was tested on a toy-model designed with experts from Orange Labs, the research division of an important commercial web actor in France, to fit the real-world problem. We took care that each advertisement campaign has its own characteristics that more or less appeal to the different visits. The model assumes that each campaign  $A_k$  has a *base click probability*  $p_k$  that is sampled from a known distribution (*e.g.* uniform in an interval, or normally distributed with a certain mean and variance). As clicking on an advertisement is in general a rare event, the base click probabilities are typically low (around  $10^{-4}$ ). The click probability of a user belonging to profile  $P_i$  is then set to  $p_{i,k} = p_k \gamma^{\mathbf{d}-1}$  where  $\gamma > 1$  is a multiplicative coefficient and the random variable  $\mathbf{d}$  is sampled from the discrete probability distribution with parameter  $n$  that has the following probability mass function  $Pr[\mathbf{d} = x] = 2^{n-x}/(2^n - 1)$ ,  $1 \leq x \leq n$ . When  $n$  is small, all campaigns will have similar click probabilities that are close to the base click probability; as  $n$  increases, some campaigns will have significantly higher click probabilities for some but not all of the user profiles<sup>1</sup>. In the experiments we used two values for the  $\gamma$  parameter, 2 and 4; experts recommended use of the latter value, but as we will see shortly having a higher  $\gamma$  value may be advantageous for the greedy policy. The value of  $n$  is varied between 2 and 6. We opted to focus on core measures and therefore omit some of the extensions that have been discussed in the text.

We begin with the static setting with full information, and consider a fixed time horizon of one day assumed to be equivalent to  $4 \times 10^6$  page visits. The distribution of user profiles is uniform and the budget and lifetime of campaigns are also sampled uniformly from fixed intervals. In order to determine the starting times of campaigns, we partitioned the time horizon into  $M$  equally spaced intervals (in our case 80) and set the starting time of each advertisement to the starting time of an interval chosen randomly such that the ending times do not exceed the fixed time horizon. The base click probability is set to  $10^{-4}$ . We solved the optimization problem every 10000 steps. Fig. 2 (a) shows the relative performance of HLP policy with respect to the HEV policy for different values of the parameter  $n$  and budget for the case in which there is a single user profile and 40 campaigns with an average lifetime of  $1/10^{th}$  of the time horizon; all campaigns have the same budget. We can make two observations, all other parameters being fixed HLP is more effective with increasing budgets, and the performance gain depends largely on the value of  $\gamma$ . For  $\gamma = 4$ , which is considered to be a realistic value by experts of the Orange Labs, and reasonable budgets the greedy policy performs well. A similar situation also arises when the number of campaigns is low, whereas increasing the number of user profiles favors planning as presented in Fig. 2 (b). Next, we consider longer static settings of over one week period with and without full information. The campaign lifetimes and their budget were more realistic (2-5 days, 500-4000 clicks). 7-9 new campaign are generated on a daily basis at the beginning of a run. We tested different values for the parameter  $n$ . There were 8 user profiles with equal visit probabilities. In this setting although HLP policy performs better than the greedy policy, the performance gain is limited (Fig. 2 (c)). While the greedy policy quickly exploits and consumes new advertisements as they arrive, HLP tends to keep a consistent and uniform click rate at the beginning and progressively becomes more greedy towards the end of the period (see [3] Fig. 10). Fig. 2 (d) shows the effect of the planning horizon; since we are not in the dynamic setting, using less information than available hinders the performance. Note that, this prominently depends on the degree of interaction between the campaigns and in this and other experiments we observed that being very far-sighted may not be necessary. Finally, we conducted experiments in the dynamic setting with partial information where the probabilities are estimated online. We employed  $\varepsilon$ -greedy exploration mechanism with different values of  $\varepsilon$  and maximum a posteriori estimation with Beta priors. The results in Fig. 2 (e) show that HLP can perform better than HEV, however for both policies the chosen hyper-parameters influence the outcome.

## 3 Related work

The oldest reference we were able to spot is Langheinrich et al. [6] who mix a linear program with a simple estimation of CTR to select advertisements to display. In this work, no attention is paid to the exploration/exploitation trade-off and more generally, the problem of the estimation of the CTR is

---

<sup>1</sup>Note that, the number of such assignments will be exponentially low; for fixed  $\gamma$ , the number of campaigns with click probability  $p$  will be twice that of with click probability  $\gamma p$ . This allows us to model situations in which a small number of campaigns end up being popular in certain user profiles.

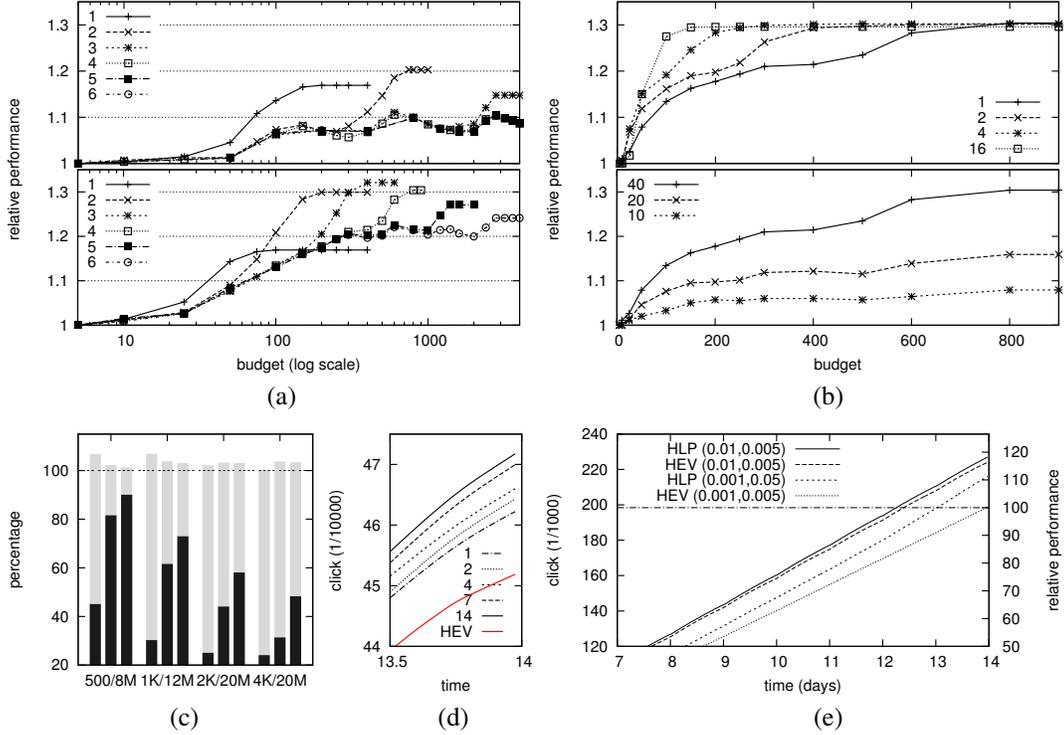


Figure 2: (a) The relative performance of the HLP policy with respect to the HEV policy for different values of  $n$  under the static setting with one profile and 40 campaigns.  $\gamma$  is 2 (bottom) and 4 (top). (b) The effect of the number of user profiles (top) and campaigns (bottom) for  $n = 2$ ,  $\gamma = 4$  and other parameters are kept constant. (c) The performance of the random (dark gray) and the HLP (light gray) policies with respect to the HES policy under the 7 days static setting for different budget (500 to 4000), lifetime (2-5 days) and  $n$  values. The three sets of bars in each group corresponds to the case where  $n = 2, 4$  and  $6$  in that order. (d) The effect of horizon (1, 2, 4, 7, 14 days) in the 14 days static setting with full information. Bottom line shows the HEV policy. (e) The performance of HEV and HLP algorithms in the dynamic setting with partial information using  $\epsilon$ -greedy exploration. The numbers in paranthesis denote the values of the parameter of the Beta prior and  $\epsilon$ .

very crudely addressed. Abe and Nakamura [1] introduced a multi-arm bandit approach to balance exploration with exploitation under unlimited resources and with a static set of advertisements. This was later improved in [11] where they address the problem of multiple advertisements on a single page, and the exploration/exploitation trade-off using Gittins indices. Ideas drawn from their work on multi-impression may be introduced in ours to deal with that issue.

Aiming at directly optimizing the advertisement selection, side information is used to improve the accuracy of prediction in several recent papers [4, 5, 8, 12, 13]. However, all these works do not consider finite budget constraints, and finite lifetime constraints, as well as the continuous creation of new advertising campaigns; they also do not consider the CTR estimation problem. Very recently, Li et al. [8] focuses on the exploration/exploitation trade-off and proposes interesting ideas that may be combined to ours (varying  $\epsilon$  in the  $\epsilon$ -greedy strategy, and taking into account the history of the displays of an advertisement). Though not dealing with advertisement selection but news selection, which implies that there is no revenue maximization, and no click budget constraint, but merely maximization of the amount click, [2, 7] investigate a multi-arm bandit approach.

A rather different approach is that of Mehta et al. [10] who treated this problem as an on-line bipartite matching problem with daily budget constraints. However, it assumed that we have no knowledge of the sequence of appearance of the profile, whereas in practice we often have a good estimate of it. Mahdian and Nazerzadeh [9] tried then to take advantage of such estimates while still maintaining a reasonable competitive ratio, in case of inaccurate estimates. Extensions to click budget

were discussed in the case of extra estimates about the click probabilities. Nevertheless, the daily maximization of the income is not equivalent to a global maximization.

## 4 Conclusion and future work

In this paper, we have provided insights on optimizing advertisement display, handling finite budgets and finite lifetimes in various settings within realistic computational time constraints. Our experimental results indicate that if there are few overlapping advertisements, or many advertisements with long lifetimes and good click rates, then we should be greedy. Between these two extreme solutions, one should consider the associated constraints. In particular, the lifetime of campaigns seem important. As future work, one possibility is to solve the problem from the perspective of the advertiser, *i.e.* help them to set the value of a click, and adjust it optimally with respect to the number of visitors (equivalent to a local sensitivity analysis of the LP problem). A more difficult issue is that of handling multiple advertisements on the same page where the correlation between the advertisements becomes important. Finally, we are also willing to draw some theoretical results on how far from the optimal strategy we are.

**Acknowledgment** This research was supported, and partially funded by Orange Labs, under externalized research contract number CRE number 46 146 063 - 8360, and by Ministry of Higher Education and Research, Nord-Pas de Calais Regional Council and FEDER through the “Contrat de Projets Etat Region (CPER) 2007-2013”, and the contract “Vendeur Virtuel Ubiquitaire” of the “Pôle de compétitivité Industries du Commerce”. Simulations were carried out using the Grid’5000 experimental testbed, an initiative from the French Ministry of Research, INRIA, CNRS and RENATER and other contributing partners.

## References

- [1] N. Abe and A. Nakamura, “Learning to Optimally Schedule Internet Banner Advertisements,” in *Proc. of the 16<sup>th</sup> ICML*, 1999, pp. 12–21.
- [2] D. Agarwal, B. Chen, and P. Elango, “Explore/exploit schemes for web content optimization,” in *Proc. of the 2009 IEEE Int’l Conf. on Data Mining (ICDM)*, 2010, pp. 661–670.
- [3] S. Girgin, J. Mary, P. Preux, and O. Nicol, “Advertising campaigns management: Should we be greedy?” Sep. 2010, INRIA Research Report RR-7388, <http://hal.inria.fr/inria-00519694/en/>.
- [4] S. M. Kakade, S. Shalev-Shwartz, and A. Tewari, “Efficient Bandit Algorithms for Online Multiclass Prediction,” in *Proc. of the 25<sup>th</sup> ICML*. New York, NY, USA: ACM, 2008, pp. 440–447.
- [5] J. Langford and T. Zhang, “The Epoch-Greedy Algorithm for Multi-armed Bandits with Side Information,” in *20<sup>th</sup> NIPS*, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds. MIT Press, 2008, pp. 817–824.
- [6] M. Langheinrich, A. Nakamura, N. Abe, T. Kamba, and Y. Koseki, “Unintrusive customization techniques for web advertising,” *Computer Networks*, vol. 31, Jan. 1999.
- [7] L. Li, W. Chu, J. Langford, and R. Schapire, “A contextual-bandit approach to personalized article recommendation,” in *Proc. of the 19<sup>th</sup> WWW*, apr 2010.
- [8] W. Li, X. Wang, R. Zhang, Y. Cui, J. Mao, and R. Jin, “Exploitation and exploration in a performance based contextual advertising system,” in *Proc. of the 16<sup>th</sup> ACM KDD*, 2010.
- [9] M. Mahdian and H. Nazerzadeh, “Allocating Online Advertisement Space with Unreliable Estimates,” in *ACM Conference on Electronic Commerce*, 2007, pp. 288–294.
- [10] A. Mehta, A. Saberi, U. Vazirani, and V. Vazirani, “Adwords and Generalized On-line Matching,” in *Proc. of the 46<sup>th</sup> Annual IEEE FOCS*. IEEE Computer Society, 2005, pp. 264–273.
- [11] A. Nakamura and N. Abe, “Improvements to the linear programming based scheduling of web advertisements,” *Electronic Commerce Research*, vol. 5, no. 1, pp. 75–98, Jan. 2005.
- [12] S. Pandey, D. Agarwal, D. Chakrabarti, and V. Josifovski, “Bandits for Taxonomies: A Model-based Approach,” in *Proc. of the 7<sup>th</sup> SIAM ICDM*, 2007.
- [13] C.-C. Wang, S. Kulkarni, and H. Poor, “Bandit Problems With Side Observations,” *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 338–355, 2005.