A Practical Exploration System for Search Advertising

Parikshit Shah
parikshit@yahoo-inc.com

Ming Yang
mingyang@yahoo-inc.com

Sachidanand Alle
alles@yahoo-inc.com

Adwait Ratnaparkhi
adwaitr@yahoo-inc.com

Ben Shahshahani
bms@yahoo-inc.com

Rohit Chandra
rc@yahoo-inc.com

ABSTRACT

In this paper, we describe an exploration system that was implemented by the search-advertising team of a prominent web-portal to address the cold ads problem. The cold ads problem refers to the situation where, when new ads are injected into the system by advertisers, the system is unable to assign an accurate quality to the ad (in our case, the click probability). As a consequence, the advertiser may suffer from low impression volumes for these cold ads, and the overall system may perform sub-optimally if the click probabilities for new ads are not learnt rapidly. We designed a new exploration system that was adapted to search advertising and the serving constraints of the system. In this paper, we define the problem, discuss the design details of the exploration system, new evaluation criteria, and present the performance metrics that were observed by us.

1 INTRODUCTION

A basic problem faced by any content delivery system is the cold-start problem: when new ("cold") content is injected into the system that competes with other "warmer" content, how does the system learn about the quality of the new content and deliver it appropriately and reliably? In addition to improved modeling techniques, a critical component is a mechanism for exploration. A typical exploration system randomly boosts cold content over existing competing content in order to learn about it. While conceptually simple, the mechanism must be designed carefully to manage the exploration-exploitation tradeoff in key metrics such as user engagement and revenue, while balancing various other implementation challenges.

The work reported in this paper was conducted with a view to improving the performance of the system on cold ads. We detail our efforts to implement and evaluate a new exploration system for the search advertisting system. While the basic exploration scheme that we have implemented is based on well-known ideas (ϵ-greedy, upper-confidence bound (UCB) based approach), we introduced novel aspects to the design that may of broader interest. Among these are:

- Methods for factoring bid information in the sampling mechanism that guards against revenue loss and provides advertisers a lever to increase the exploration rate for their cold ads,
- New practical methods of evaluation - such as a novel way of tracking the performance of good versus bad ads found via exploration. Since our exploration system is tightly coupled with the production click model, and the design and evaluation are crucially influenced by the same; we describe this interaction in later sections.

Exploration has a rich history in the machine learning literature [2–5, 9]. Exploration approaches have been studied in various contexts such as news article recommendation, native advertising, search advertising [9, 10], reinforcement learning and its application to computer games [8]. Furthermore, exploration has also been used to radically improve the traditional A/B-testing paradigm such as the work on Multi-World Testing [1]. In contrast to previous existing works in the applied literature, which focus on designing exploration to achieve fast online learning with low regret, in this work we describe a practical approach that is specifically tailored to search ads and is thus tailored to the ad auction – both in terms of the exploration implementation as well as the evaluation.

This paper is organized as follows: In Section 2 we discuss the relevant background concerning the search advertising system. In Section 3 we describe and discuss the details of our exploration system. In Section 4 we discuss the evaluation criteria and the metrics.

2 BACKGROUND

Search advertising, the method of placing relevant online advertisements on web pages that show results from search engine queries, has become an important part of the online user experience. Search advertising is an extremely attractive proposition for advertisers because the search query provides a powerful relevance signal that can be used for targeting only the most appropriate ads. Conceptually, a typical search advertising system consists of the following components:

- Campaign management system: this is an inventory of advertisements (or creatives), along with their title, description, and search keywords based on which they become eligible for display when the search query is similar to the keyword. Each advertiser creative also has associated with it a bid. Importantly, new advertisments are introduced by the adverisers in the campagin management system periodically.
- Matching: The matching system is reponsible for understanding the query and retrieving all the relevant ads (from the campaign management system) that match the query context.

KDD'17, August 13–17, 2017, Halifax, NS, Canada.

DOI: http://dx.doi.org/10.1145/3097983.3098041

- Click-model: The click-model reports an estimate of the probability of a click for ad a, in the context of a query q by a user u, denoted by $p_{q,a,u}$. Each element of the triple (q, a, u) can be further expanded into more detailed features; for example the query can be broken down into the tokens and the characters, the advertisement is associated to a particular doman, campaign, ad-group and has title and description features, and lastly the user information may contain the user-cookie, location, time, gender, IP address type, etc. which are used as training inputs to the model.
- Auction: The auction determines the final selection of creatives to be displayed, their relative page positions, and the price-per-click for each. Since search revenue is click-driven (i.e. an ad is only monetized if upon display it is clicked by the user), both the bid as well as the click probability is used in a rank-score for selection/ranking of final ads. The most commonly used rank-score is the product of the bid and the estimated click-probability.

A basic challenge that is faced in the context of search advertising is the *cold-ad* problem. A cold ad refers to a creative that is introduced by an advertiser which has relatively few or no historical impressions (we call the latter "frozen" ads). The cold ad problem is actually a two-fold problem:

(1) When the ad is cold and has insubstantial historical impressions, many of the ad-specific features are unknown and thus inaccurately learnt. Thus, the predicted CTR (i.e the click-through-rate) of the ad, i.e. $p_{q,a,u}$ is not accurate and can affect the performance of this creative in adverse ways. This in turn can cause the overall system to behave suboptially, either showing poor quality new ads or showing stale ones repeatedly instead of new, better quality ads.
(2) When a new ad is introduced by an advertiser, the intent is for it to gain impression volume. However, due to the predicted CTR being inaccurate, some new ads may be underpredicted and fail to gain impression volume.

2.1 Impression Volume and Colds Ads

From the advertiser's perspective, the most serious consequence is a lack of impression volume of the cold ad. In a typical scenario, an advertiser introduces a new creative as part of an existing campaign, or a new campaign altogether. However, the creative in question fails to get impression volume for one or more of the following reasons:

- Ad-Quality Filtering (AQF): In order to maintain a certain minimum quality of content displayed online, ads deemed low-quality (in the context of a query) are often filtered out and made ineligible for the auction. Thus, regardless of the bid on such a creative, it may be filtered if its quality score is too low. The primary quality score currently used within the system is the predicted click probability $p_{q,a,u}$. Thus, if for a cold ad $p_{q,a,u}$ is below a threshold t_{AQF}, the cold ad will be blocked from gaining impression volume (for that query and user combination).
- Reserve Price: Regardless of the presence of other competing ads, every advertisement must clear a particular reserve price in the auction in order to obtain impressions, i.e. the product of its bid and predicted CTR must be above a certain threshold (we denote it by t_{reserve}. If the creative falls below this monetization threshold, the creative is deemed to be not worth (in monetary terms) the adverse user experience created by the introduction of the ad on the search page.
- Auction: A creative can be blocked from gaining impression volume because it competes in a deep-market query with many other ads with a higher rank-score. For instance, if no more than five ads are eligible for display on the search page, and there are at least five other ads with a higher score (i.e. the product of bid and predicted CTR), the new creative will not gain impression volume.

A crucial observation is that the above blockers to impression volume are all sensitive to the predicted CTR, $p_{q,a,u}$. AQF directly depends on the predicted CTR, and the CTR is the only system-related factor in the reserve price as well as the auction (since the bids are controlled by the advertisers). Thus, if the CTR is predicted inaccurately on cold ads, impression volume is directly impacted. Furthermore, a substantial volume of creatives returned by matching are cold, and thus the overall performance of the system depends crucially on predicting on cold ads accurately. Fig. 1 shows a histogram of impression volume by its coldness level, and indicates that a significant fraction of ads returned are cold/frozen (a frozen ad is one that has no impression history whereas a cold ad is one with small impression history).

2.2 Click Model and Cold Ads

The current production click-model is a supervised feature-based model:

$$p(\text{click}|q, a, u) = p_{q,a,u} := F(f_j(q, a, u))$$

where $\left\{f_j(q, a, u)\right\}_{j=1,\ldots,N}$ is the j^{th} feature extracted for query q, ad a and user u, and F is chosen from a parameteric function class (such as logistic regression or gradient boosted decision trees) and the parameters are selected by empirical loss minimization from historical data [6]. The most important features of the model (i.e. the ones that largely determine the predicted click-probability) are the so called click-feedback features, also known as the EC/COEC features [6, 7].

- EC features: These capture position-normalized impressions for the corresponding feature. The EC feature for a creative, denoted *CRTEC* is especially important for exploration since it counts the number of times (suitably normalized) the ad has been shown historically, i.e. a measure of how cold the ad is.
- COEC features: These capture click information. The CRT-COEC is defined to be the number of historical clicks achieved by the creative id in question normalized by the EC, and is a proxy for the position-normalized historical CTR for that ad.

During training, the feature weights are learnt for each feature of the form described above. Accurate click prediction depends crucially on the correct learning of feature weights. The feature weights for features that have occured very few times in training are quite unreliable. For instance, a creative that has only 5 impressions

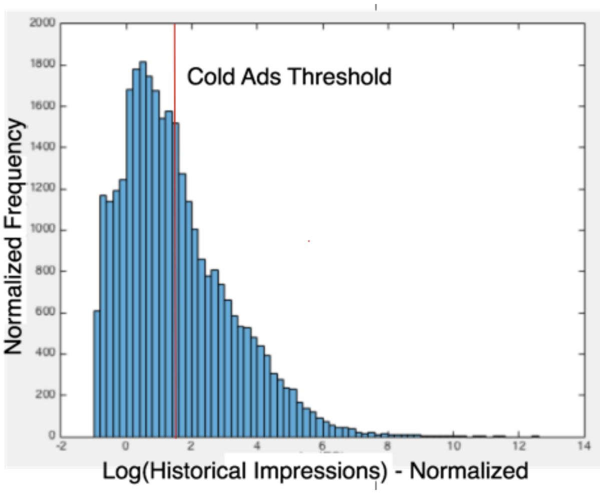

Figure 1: We show a histogram of the number of historical impressions of ads that are returned by the ad-matching system on a particular day of traffic. Note that the x-axis has the number of historical impressions corresponding to ads, and these are normalized for position effects, and also rescaled so as not to reveal sensitive proprietary data. The y-axis correspond to the frequency with which ads falling into the corresponding historical count bucket appear on that day. Ads to the left of the red line are considered cold ads. We see that a substantial number of ads returned are cold (i.e. the histogram is skewed to the left), thereby emphasizing the scope of the problem.

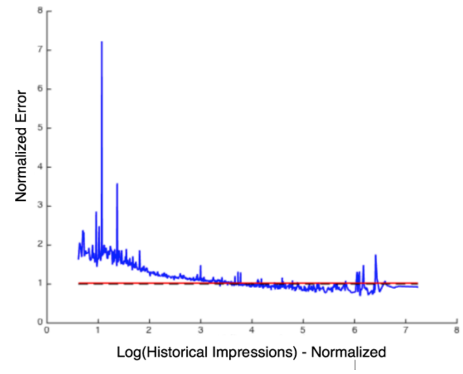

Figure 2: We show how the click-model's accuracy behaves as a function of the historical impressions recieved by corresponding creative. On the x-axis is the logarithm of the number of historical impressions (normalized for position effects, and rescaled), and the y-axis is a notion of error called the relative bias of the ad. We see that for cold ads, the error is substantially worse than for warm ads.

with a single click will learn a CRTCOEC weight such that the predicted CTR would be close to 0.2. However, if the click were accidental, and the true CTR was much closer to .02, the learnt weight would be substantially incorrect. An even more extreme case is when the CRT in question has never been seen before (in which case the feature takes on a default value and the weight corresponds to the historical weight for all frozen CRTs). In this situation, if the CTR for the frozen ad is underpredicted, not only will the CRT not gain impression volume (adverse for advertiser), but as consequence the EC and COEC values will continue to remain incorrect and click model will never have an opportunity to learn the true CTR. Fig. 2 shows the click-prediction accuracy of ads as a function of the number of historical impressions of the creative. The figure clearly shows that predictions for cold ads are more inaccurate as compared to warm ads. Exploration is a crucial component for the long-term health of a search advertisting system. Unless cold ads are impressed, the system cannot learn its features, and there is a danger that it would keep showing stale ads. This may be bad not only for advertisers seeking volume for new, cold ads, but also suboptimal in terms of revenue and user experience for the overall system.

2.3 Exploration-Exploitation Tradeoff

Intimately tied to the cold ads problem is the well-known *exploration-exploitation* tradeoff [10]. In this context the tradeoff is clear: any given query yields multiple ads competing in the auction with different rank-scores. The warm ads have a lower variance in the predicted CTR (and hence a lower variance in the rank-score). The cold ads, have a higher variance. Exploitation would correspond to picking the safe option, i.e. warm ads with highest rank-scores. Exploration would correspond to picking the cold ads with none or few historical impressions, thereby discovering potentially new high-quality ads at the risk of also showing low-quality ads and thereby losing revenue. Understanding this tradeoff has a rich literature in machine-learning [2–5, 9].

Problems of this flavor are often studied as *multi-armed bandit* problems [3]. In this setting, the decision-maker is presented with the opportunity to select (or "pull") one (or more generally k) arms from a selection of n arms of a slot machine. Each arm (in our context, an arm is an ad) has a random reward with a fixed but unknown distribution. The objective of the decision-maker is to, over multiple rounds, maximize her expected reward. Once an arm is pulled (i.e. ad is shown), a reward from the distribution of the arm is realized, and the decision-maker may learn from this information. A pertinent modification of the multi-armed bandit problem is the *contextual* bandit problem. In this setting, each arm also presents a context or a feature vector. In our setting, the feature vector would correspond to the click-modeling features (i.e. information about the query, ad, and user).

Popular methods for exploration-exploitation involve sequentially acting and learning statistics of the unknown arm distributions. For each arm, a prior distribution is assumed. At each stage, an action is performed (selection of arm i), a reward is obtained, and the posterior distribution (or some sufficient statistics) corresponding to the pulled arm i is updated. Most commonly, the prior is assumed to be normal and the main statistics that are tracked are the mean of the arm (i.e. $\hat{p}_i$ as well as the variance of the arms $\hat{\sigma}_i^2$). Some well-known approaches for the same [3, 4] are the following:

- Thompson sampling: in this approach, the rewards are sampled from the posterior distribution and the arm with the maximum sampled reward is picked.
- UCB: in this approach the decision rule is to pick the arm with the highest expected reward plus a standard deviation (i.e. the upper confidence bound).
- ϵ-greedy: In this approach, exploration is only performed on a fraction ϵ of the rounds, either completely randomly, or as per one of the two above strategies.

We adopted UCB over the Thompson sampling based approach as we favor a deterministic exploration approach as opposed to a randomized one (sampling from posterior), both for computational reasons as well as ease of post-hoc analysis. In order to be conservative, we combined UCB with an ϵ-greedy approach where only fraction of query sessions were made available for exploration (the other session were pure exploitation).

3 A PRACTICAL IMPLEMENTATION

In our implementation we adopt the UCB based exploration strategy with an ϵ-greedy exploration. The precise algorithm is described in Algorithm 1.

Algorithm 1 Exploration Algorithm for Search Ads

Input: query q, user u, ads $a_1, \dots a_m$ their corresponding click probabilities $p_1, \dots, p_m$, their bids $b_1, \dots, b_m$, and their historical (position-normalized) impression counts $n_1, \dots, n_m$, blacklist of ads B.

Parameters: ϵ, the greed parameter; eligibility thresholds $p_{\text{th}}, n_{\text{th}}$ for the probability and coldness; $r_{\max}$ the maximum number of arms that can be pulled in a single round of exploration; the boosting parameters α, β.

ϵ-Greedy: Sample from Bernoulli distribution $X \sim B(\epsilon)$. If $X = 0$, send $F = \{(a_i, p_i) | i = 1, \dots, m\}$ to auction and **return**.

Let $E = \emptyset$

Eligibility: For each i, if $p_i < p_{\text{th}}$ and $n_i < n_{\text{th}}$ and $a_i \notin B$, add $E = E \cup \{i\}$.

Bid-proportional sampling: From E sample $q = \min(|E|, r_{\max})$ ads randomly with replacement with probability $\theta_j = \frac{b_j}{\sum_{k=1}^{|E|} b_k}$. Let the selected ads have indices $T = \left\{i_1, \dots, i_q\right\}$.

Boosting: For each $j \in T$, $\bar{p}_j := p_j \left(1 + \frac{\alpha}{\sqrt{\beta + n_j}}\right)$.

Form final set: Define

$$F := \bigcup_{j \in T} \left\{(a_j, \bar{p}_j)\right\} \cup \bigcup_{k \in T \setminus [m]} \{(a_k, p_k)\}$$

Output: Return F to the auction.

We discuss a few implementation-related points worthy of mention:

- **ϵ-greedy:** We pick an ϵ-greedy strategy, wherein only an $\epsilon = .05$ fraction of query sessions are (randomly) chosen for exploration. The principal reason is to be conservative, and perform exploration in a low-risk, revenue neutral manner. Furthermore, the impression volume is large enough that with $\epsilon = .05$, an effective amount of exploration is achieved.
- **Eligibility:** When a query session is activated for exploration, a number of ads are returned by matching for consideration. Not all of these ads are made eligible for exploration. Ads that have sufficient historical impression volume (i.e. warm ads) have a reliable predicted CTR, and there is no need to explore these ads. We choose the threshold n_{th}, by plotting the accuracy as a function of historical impressions. The value is chosen to be $n_{\text{th}} = 500$. We also do not explore cold ads whose predicted CTR is under a certain threshold. The value is chosen to be the one corresponding to the AQF module ($p_{\text{th}} = .02$), i.e. all ads below which are filtered by the AQF module. Lastly, we also maintain a blacklist B of ads that are explicity prohibited from exploration. This blacklist corresponds to known undesirable ads that cause a poor user experience.
- **Boosting:** The main mechanism that facilitates exploration is boosting of the predicted click probability. The formula for boosting is inspired by the UCB algorithm. Intuititvely, the variance of the predicted click probability is assumed to behave as $\alpha \frac{p(1-p)}{n}$, where n is the number of samples. Assuming $1 - p \approx 1$, the upper confidence bound assumes the above form. In the formula, α is a parameter to be tuned, it was picked to ensure that more than 80% of explored ads would clear the AQF threshold. The quantity β is also a tuning parameter and is in place to ensure that when the number of impressions is zero, the boost factor remains bounded. Boosting-based exploration of the above form has been studied in the literature [9].

  Note that only some of the cold ads are boosted. The warm ads are never boosted and this amounts to the assumption that the uncertainty in the estimate is essentially zero. Note also that the above approach of boosting is contextual because it uses the predicted click probability p_j in the boost function, which in turn depends on the feature vector $f(q, a_j, u)$.
- **Auction:** We note that a significant difference between exploration in search advertising as opposed to the conventional multi-armed bandit problems (e.g. recommending news articles, see [9]) is the presence of the auction. In our context, pulling an arm is tantamount to boosting the predicted CTR for the corresponding ad. Post-boosting, the ad will still compete in an auction, and will translate into an impression only when its rank-score (product of bid and predicted CTR) is high enough to win in the auction. Thus, unlike the traditional setting, "pulling an arm" (boosting), is not guaranteed to produce an impression and thus does not necessarily provide information from which the click model can learn. As we will see later, measuring whether the click model actually learns is an important success criterion for the exploration design.
- **Bid-proportional sampling:** A large number of ads may be returned by matching for a given query session. Recall that from a bandit viewpoint, the ads returned are the arms. Since only a small number of ads may be displayed

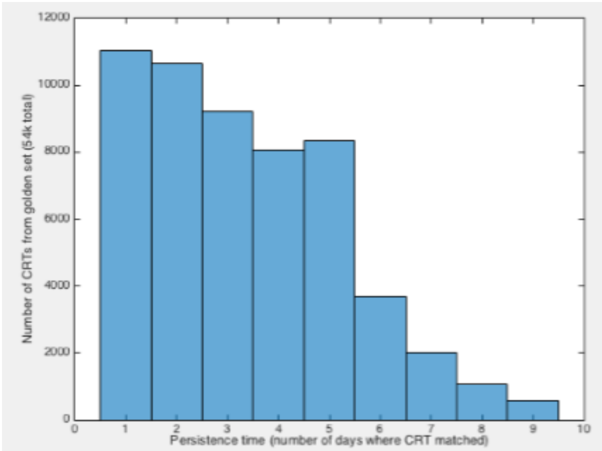

Figure 3: This plot is a histogram for the persistence of the ads returned in a 9 day experiment. Each bar shows the number of cold ads (in bucket) that were returned by matching at least once on the corresponding (x-axis) number of days. The plot demonstrates that a large fraction of cold ads are not persistent, i.e. they appear only once in a nine-day time-window, and do not appear again. Such non-persistent ads cannot be learnt, and are ignored from the subsequent learning rate analysis.

on a page, boosting too many ads may displace all of the legitimate high-quality ads the exploration and thereby hurt revenue/user-experience. The maximum number of arms that may be pulled $r_{\max}$ is fixed up-front. We set $r_{\max} = 2$.

In certain situations, the number of ads that are eligbile for exploration may be smaller than $r_{\max}$, in which case we decide to explore all the eligible ads. However, in most situations, the number of eligible ads returned is larger, and we therefore must decide which ads to explore (i.e. boost). While randomly choosing (with equal probability) is a natural approach, we found sampling proportional to the bid to be quite effective. This approach has multiple advantages. The first advantage is that by boosting high bid ads, we increase the likelihood that the boosted ads win a page slot in the auction. The second advantage is that biasing the sampling toward the high bid ads keeps the price-per-click (PPC) of the auction high, thereby guarding against a revenue loss. Third, this provides a lever for advertisers with cold ads in the system to increase their volume: they can get higher impression share by simply increasing their bids.

- **Logging/Learning:** An important part of the platform is the data-logging, a detailed record of every impression and click is kept in a data feed. Due to serving latency constraints, only the impressed ads are logged. Importantly, these impressions become part of both the click feedback EC/COEC features (which are updated) and also the training data for the next iteration of model update (this is essentially how learning is achieved from exploration). Ads that competed in auctions for pageviews, but were not impressed are not logged. Understanding the performance of exploration critically relies on these competing ads. Hence we enable a special logging mechanism in the adserver for the exploration bucket that logged all ads that were returned by the matching module to the click model. Information of this nature enabled us to gain insights such as Fig. 1 which were helpful in designing and tuning the exploration parameters.
- **Persistence:** An interesting observation related to our experiments was that while there was found to be a high volume of cold ads, a number of these were not persistent, i.e. these ads were not consistently returned by matching (see Fig. 3) for competing in the auction (over multiple instances of the same query). This type of behaviour can occur due to budget constraints imposed by the advertisers on their campaigns, or the inherent randomness in the matching algorithms themselves. Learning click probabilities for ads that are not persistent is much harder, and limits the efficacy of exploration. It is desirable to focus exploration only on persistent ads. We found that bid-proportional sampling helped bias the exploration toward persistent ads.

4 EVALUATION

We implemented the design of the exploration scheme described in the preceding section and ran a bucket experiment. The test bucket was instrumented with a special form of logging; for each query session not only were the impressions logged, but also all the ads that were returned by matching but not impressed (along with relevant information such as their click probabilities, boost factors, bids, and other relevant information). The experiment was successfully concluded after a few weeks of experimentation.

The evaluation process involved understanding three different kinds of metrics. The first correspond to understanding the effect of exploration on the click model features (thereby influencing the accuracy of the model on cold ads); we refer to this as the learning rate. The second involved understanding the impact on the usual business metrics. Finally, we also investigated whether or not the ads explored were "good" versus "bad" — we concretely define these notions subsequently. We discuss these evaluations in detail in the sections below.

4.1 Learning Rate Metrics

A widely used metric that characterizes learning achieved by exploration is the notion of regret [3]. In the regret framework, the problem of learning click probabilities may be viewed as an online learning problem where some loss function (e.g. revenue or clicks) must be optimized in an online manner. The performance of any given online algorithm may be compared with the performance of a hypothetical decision maker with perfect hindsight information. The difference between the optimum achieved between the two is called the regret (which can be analytically characterized), and quality of an algorithm is characterized by how low its regret is.

A basic drawback of regret as a solution-concept is that it is quintessentially counterfactual in nature. Since the hypothetical decision-maker can never be realized, the regret cannot be explicitly measured in a bucket experiment, and it is difficult to justify from a business standpoint. Hence we seek an alternate solution concept that is more closely tied to the product viewpoint.

To arrive at our notion of "learning" we recognize that the click-model uses logistic regression on a set of features. To predict accurately, the weights must be correctly identified. The most important feature that determines the accuracy of the prediction is the CRTEC, i.e. the number of times a particular creative has been seen historically (i.e. on the training data), see Fig. 2. Hence, to predict accurately for a given cold ad, exploration must achieve enough impressions so that its click probability is accurately learnt.

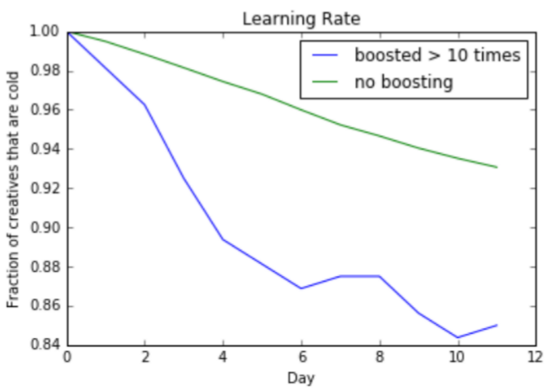

Figure 4: This plot shows the learning rate achieved for a pool of cold, persistent ads over an 11 day experiment. It shows that boosted ads become warm at a faster rate, hence resulting in faster learning.

More concretely let C be a set of cold ads (represented as creative ids), fixed on day 0 of the experiment. Let $c \in C$ be a fixed creative, and let $n_c(t)$ be the number of impressions of c on day t of the experiment. Let N_0 denote the coldness threshold, i.e. the threshold above which ads are considered warm. The fraction of creatives that remain cold at day t is then defined as $f(t) := \frac{\sum_{c \in C} \mathbf{1}(n_c(t) < N_0)}{|C|}$, where $\mathbf{1}(\cdot)$ denotes the indicator function. We will use $f(t)$ as the learning rate metric to track. A faster decay of $f(t)$ indicates that more ads have become warm (hence the click prediction should be more accurate). In Fig. 4 for a fixed set of creatives, we show the learning rate plot with and without exploration.

4.2 Business Metrics

The standard business metrics of interest are mentioned below. We explain the role of each metric, along with the bucket metric for traffic on tablet devices (performance on desktop was found to be directionally similar). All results presented below are obtained from bucket experiments conducted over multiple weeks with sufficient data so that the results are traditionally regarded as statistically significant (p-value below 0.001).

- Revenue-Per-Mille (RPM): This is the revenue generated per thousand impressions. We found that the RPM was +1% as compared to the control bucket.
- Price-Per-Click (PPC): This is the price per click as determined by the auction. The PPC was +0.5% as compared to the control bucket
- Click-Through Rate (CTR): This is the number of clicks per unit of impression (normalized by position) in the north section of the page. The CTR was found to be +.04% as compared to control.
- Click Yield (CY): The click yield is the number of ad clicks per total number of searches. The click yield was found to be +0.5% compared to control.
- North footprint (NFP). This correponds to the number of north ads shown per total number of searches. The NFP was found to be +0.2%. The objective is to achieve positive metrics for all the above quantities while keeping NFP as neutral as possible.

Our exploration design was able to demonstrate learning (in the sense explained above) while maintaining the above revenue and user engagement metrics neutral, and the experiment was thus concluded successfully.

4.3 Good versus Bad Ads

In addition to the learning-rate metrics and the business metrics that were used for the evaluation of the bucket experiments, the product team also wished to understand whether exploration was above to discover new ads of high-quality. We hence evaluated this question by investigating the set of ads that were explored over a larger volume of traffic.

To define this notion more formally, we fix a cold ad a, on day 0 of the evaluation. Let $p_{a,\mathrm{init}}$ be the predicted CTR for the ad a (aggregated across all the different queries and users in which the ad is returned by matching on day 0). On day 0, because the ad a is cold, the empirical CTR of the ad is unknown. Assuming the ad is boosted and impressed a sufficient number of times in the next N days, the ad a will become warm, and its empirical CTR $p_{a,\mathrm{emp}}$ is known

For the purposes of exploration, we say that a cold ad is "good" when $p_{a,\mathrm{emp}} - p_{a,\mathrm{init}} > \delta$. (In other words the CTR of the explored ad is higher than the predicted click probability of the ad pre-exploration, and hence exploration was able to discover a good ad). Conversely, we say that an ad is bad when $p_{a,\mathrm{emp}} - p_{a,\mathrm{init}} < -\delta$. We set $\delta = .001$, and find that in 20 days of exploration data, for every creative that was persistent (appeared at least 100 times over multipe days), the following metrics:

Good Ads	Bad Ads	Neutral Ads
9.5%	85.9%	4.6%

5 CONCLUSION

In this paper, we described our efforts to build a production-scale exploration mechanism for a search advertising system. We described various pitfalls encountered in the process, novel evaluation criteria, and important metrics. Specifically, we found that, the exploration system explored a number of new ads, of which 9.5% were found to be good - i.e. these were ads with higher than expected click-through-rate, which would have been otherwise discarded by the advertising system. While we argued, that the notion of regret (which is typically studied in the machine learning theory community), due to its' post-hoc nature, is difficult to measure, we proposed tracking good vs. bad ads as a novel, useful, and practical measure of evaluation for an exploration system.

REFERENCES

[1] A. Agarwal, S. Bird, M. Cozowicz, L. Hoang, J. Langford, S. Lee, J. Li, D. Melamed, G. Oshri, O. Ribas, S. Sen, and A. Slivkins. A multiworld testing decision service. *CoRR*, abs/1606.03966, 2016.

[2] D. Agarwal et al. Online models for content optimization. In *Advances in Neural Information Processing Systems 21*, pages 17–24. Curran Associates, Inc., 2009.

[3] S. Agrawal and N. Goyal. Thompson sampling for contextual bandits with linear payoffs. In *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013*, pages 127–135, 2013.

[4] P. Auer. Using confidence bounds for exploitation-exploration trade-offs. *J. Mach. Learn. Res.*, 3:397–422, Mar. 2003.

[5] O. Chapelle and L. Li. An empirical evaluation of thompson sampling. In *Advances in Neural Information Processing Systems 24*, pages 2249–2257. Curran Associates, Inc., 2011.

[6] O. Chapelle and Y. Zhang. A dynamic bayesian network click model for web search ranking. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 1–10, 2009.

[7] H. Cheng and E. Cantú-Paz. Personalized click prediction in sponsored search. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, pages 351–360, New York, NY, USA, 2010. ACM.

[8] D. Silver et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 01 2016.

[9] L. Li, W. Chu, J. Langford, and R. E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 661–670. ACM, 2010.

[10] W. Li, X. Wang, R. Zhang, Y. Cui, J. Mao, and R. Jin. Exploitation and exploration in a performance based contextual advertising system. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, pages 27–36, New York, NY, USA, 2010. ACM.