Fast and Cost-efficient Bid Estimation for Contextual Ads

Ye Chen, Pavel Berkhin, Jie Li, Sharon Wan, Tak W. Yan
Microsoft Corporation
1065 La Avenida, Mountain View, CA 94043
{yec,pavelbe,lijie,xiawa,takyan}@microsoft.com

ABSTRACT

We study the problem of estimating the value of a contextual ad impression, and based upon which an ad network bids on an exchange. The ad impression opportunity would materialize into revenue only if the ad network wins the impression and a user clicks on the ads, both as a rare event especially in an open exchange for contextual ads. Given a low revenue expectation and the elusive nature of predicting weak-signal click-through rates, the computational cost incurred by bid estimation shall be cautiously justified. We developed and deployed a novel impression valuation model, which is expected to reduce the computational cost by 95% and hence more than double the profit. Our approach is highly economized through a fast implementation of kNN regression that primarily leverages low-dimensional sell-side data (user and publisher). We also address the cold-start problem or the exploration vs. exploitation requirement by Bayesian smoothing using a beta prior, and adapt to the temporal dynamics using an autoregressive model.

Categories and Subject Descriptors

I.5.1 [**Models**]: Statistical

General Terms

Algorithms, Theory, Experimentation

Keywords

Computational advertising, nonparametric models, exchange

1. INTRODUCTION

We study the problem of estimating the value of a contextual ad impression, with the goal of returning a sufficiently accurate value-based bid to an ad exchange in a cost-efficient manner to optimize profit or return on investment (ROI) for an ad network. An ad exchange sends a request for bid (RFB) for an ad impression to an ad network or more generally a bidder. The bidder then returns a bid based upon the value of the impression or the revenue opportunity through winning the auction and user clicking on the served ads. The value-based bid estimation or simply impression valuation incurs the computational cost-to-bid (CTB). If the bidder wins the impression, she pays the next highest bid (second-price auction) as the traffic acquisition cost (TAC). The ad network then runs an internal generalized second-price auction (GSP) to select which advertisers' ads to serve, along with their ranking and pricing, in response to the request for payload (RFP) from the exchange and incurring the computational cost-to-serve (CTS). If the served ads are clicked, the advertisers pay the ad network the click-through rate (CTR) adjusted next highest cost-per-click (CPC) bid, as in standard GSP pricing [1]. We will focus on impression valuation in response to a RFB in this work.

Impression valuation plays a central role in bidding for performance-based ads, yet is a difficult task involving estimating rates of very rare events, i.e., click-through of contextual ads. Given a contextual ad call impression or block i, the expected revenue is

$$\mathbb{E}(y_i) = w_i \sum_{k=1}^{m} b_{i,k+1} p_{i,k+1}, \tag{1}$$

where w_i is the auction winning rate, k indexes m positional ranks contained in the ad impression block, $b_{i,k+1}$ and $p_{i,k+1}$ are the CPC bid and the estimated CTR, respectively, of the ad ranked $k+1$ via a GSP. The expected payoff is then

$$\mathbb{E}(u_i) = w_i \left(\sum_{k=1}^{m} b_{i,k+1} p_{i,k+1} - c_i^{\text{TAC}} - c_i^{\text{CTS}} \right) - c_i^{\text{CTB}}, \tag{2}$$

where c_i^{TAC}, c_i^{CTS} and c_i^{CTB} are the TAC, CTS and CTB terms, respectively. CTB is primarily incurred by bid estimation and would sink regardless of winning an auction or not (w_i), nor depends upon the CTR ($p_{i,k}$); hence more difficult to justify under a competitive exchange for low-response impression opportunities. A typical value of w_i in an open exchange is 10%, and $p_{i,k}$'s of contextual ads are very low, e.g., 0.3%. A full-fledged valuation would scan hundreds of campaigns for a best match to realize the impression value thus costly in terms of CTB; while only relying on the sell-side data of the impression, i.e., user and publisher, may already give an accurate enough estimate with a substantially lower CTB.

2. AN EFFICIENT kNN REGRESSION

The expected revenue given a won impression is referred to as the true value of the impression, as shown in Eq. 1 excluding the winning probability. We wish to estimate the true value using a predictive model of the general form $y = f(x)$. This is a regression problem involving two stochastic processes: (1) a GSP mechanism, and (2) the click-through thereupon. With economical computation as one design goal, x is an input vector encoding only the sell-side features of the impression, which are known and unique from the RFB at run time, e.g., user geolocation, publisher, page URL, and ad placement. We use a k-nearest-neighbor (kNN) regression to memorize an aggregated view of history, to implicitly capture the best match with the buy-side data, e.g., advertiser and ad. Formally, given a dataset of historical impression valuation $D = \{x_i, y_i\}_{i=1}^{n}$,

WWW 2012 Companion, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1230-1/12/04.

the offline training involves building a mapping:

$$f : x \to (n_x = \sum_{i:x_i=x} 1, s_x = \sum_{i:x_i=x} y_i), \tag{3}$$

where x denotes a point in feature space. In online prediction, given an ad impression x', the MLE of the value is given by the first moment

$$y^{\text{MLE}} \leftarrow \begin{cases} s_x/n_x & \text{if } x' = x, \\ s_{z(x)}/n_{z(x)} & \text{otherwise.} \end{cases} \tag{4}$$

Here $z(x)$ is a mapping function from x to an aggregated level z, e.g., the publisher-placement pair, n_z and s_z are accumulated accordingly.

Since kNN is a nonparametric model, it will bravely predict zero for x without any historical clicks. This behavior is desirable when sufficient impressions have been seen. However, at the beginning of launching a model on a new traffic source, e.g., a new page, some form of exploration vs. exploitation needs to be built-in. One way to approach this is to impose a beta prior on y, derived from an aggregated level z naturally available from domain hierarchy and typically with much denser data, as follows

$$y_x \sim \text{Beta}(\lambda y_{z(x)} + 1, \lambda(1 - y_{z(x)}) + 1), \tag{5}$$

where $y_z = s_z/n_z$ and λ is the smoothing factor. The MAP estimate of the value of impression x is

$$y_x^{\text{MAP}} \leftarrow \frac{s_x + \lambda y_{z(x)}}{n_x + \lambda}. \tag{6}$$

The Bayesian interpretation is that we have *a priori* observed $\lambda y_{z(x)}$ revenue from λ impressions with feature vector x before we see any real x. λ controls the smoothing strength, and we wish to have a reasonably strong smoothing for those x's with zero revenue $s_x = 0$, while being conservative with the x's with sufficient positive feedbacks, especially for $s_x \gg 0$. One data-driven approach is $\lambda \leftarrow \text{mode}(n_x : s_x = 0)$. This ensures that for most zero-revenue x's, the MAP estimate is half its back-off estimate $y_{z(x)}$.

The kNN estimator derived thus far assumes that the expected value y_x stays static temporarily. In practice, however, the system is dynamic, especially in an exchange environment due to supply (inventory mix) or demand (user and campaign concept drift) changes. To adapt to the temporal dynamics, we apply an autoregessive model to decay the importance of old data, as follows

$$y_x^{\text{Dyn}} \leftarrow \frac{\sum_{t=1}^{T} (s_x^t + \lambda y_{z(x)}^t) \exp(\gamma t)}{\sum_{t=1}^{T} (n_x^t + \lambda) \exp(\gamma t)}, \tag{7}$$

where t indexes $1 : T$ training days, γ is an exponential decay parameter fitted into the latest training day T using least squares and updated daily. The existence of temporal dynamics and the effectiveness of our approach are quite evident empirically, as shown in Figure 1.

We now comment on the rationale behind our choice of kNN. Clicking on contextual ads is not only a very rare event, with more than 95% ads getting no response; but also a very random event, with about 90% variance of revenue cannot be explained by any single feature available. It is known that kNN classifier is universally Bayes-consistent under the following sufficient condition [3]: if $n \to \infty$, then $k \to \infty$ and $k/n \to 0$. Our implementation approaches this condition by controlling the feature dimentionality. Most features in ad domain are categorical, and we use binary encoding, i.e., each feature value is a dimension. For each feature, we first select values by document frequency and use a minority bin to hold rare ones. By such feature value selection, we ensure a desired

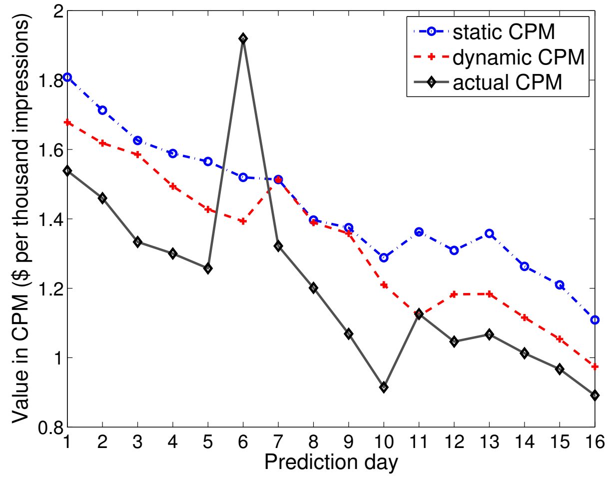

Figure 1: Static and dynamic prediction vs. actual.

overall dimensionality d, and the number of kNN keys is upper bounded by 2^d. Consequently, if $n \to \infty$, then $k \approx n/2^d \to \infty$ and $k/n = 1/2^d \to 0$. Empirically, we have tried linear regression and decision tree, and both yielded suboptimal results.

3. EXPERIMENTAL RESULTS

We conducted a two-week period online A/B testing on MSN web traffic, to compare the kNN valuation model with the current production model, which is a probit regression full-evaluation model [2]. The evaluation metrics are primarily (1) sum of prediction error rate: $\sum_i (\hat{y}_i - y_i)/\sum_i y_i$, and (2) profit as in Eq. 2. The results are reported in dollars per thousand impressions (CPM) as shown in Table 1.

Table 1: Online Testing Results

Model	kNN		Probit regression	
Test week	Week 1	Week 2	Week 1	Week 2
Win rate	12%	7.7%	9%	4%
Actual CPM	1.35	1.00	1.43	1.65
Predicted CPM	1.07	1.15	1.32	1.09
Error rate	−21%	15.30%	−8%	−34%
TAC and CTS	0.13	0.12	0.14	0.15
CTB	0.03	0.05	0.89	2.00
Profit	1.18	0.83	0.40	−0.50

The results show that kNN reduces the computational cost of impression valuation (CTB) by 95%, hence yields more than twice profit, compared with the current model. We also observe that kNN tends to have higher auction winning rate, which suggests that the nonpamametric approach yields better calibrated estimates and in turn further increases ROI.

4. REFERENCES

[1] B. Edelman, M. Ostrovsky, and M. Schwarz. Internet advertising and the generalized second price auction: Selling billions of dollars worth in keywords. *American Economic Review*, 97:242–259, 2007.

[2] T. Graepel, J. Q. Candela, T. Borchert, and R. Herbrich. Web-scale Bayesian click-through rate prediction for sponsored search advertising in Microsoft's Bing search engine. *ICML 2010*.

[3] C. J. Stone. Consistent nonparametric regression (with discussion). *Annals of Statistics*, 5:595–645, 1977.