Traffic Quality based Pricing in Paid Search using Two-Stage Regression

Rouben Amirbekian, Ye Chen, Alan Lu, Tak W. Yan, Liangzhong Yin
Microsoft Corporation, 1020 Enterprise Way, Sunnyvale, CA 94089
{roamirbe,yec,alanl,takyan,liangzhong.yin}@microsoft.com

ABSTRACT

While the cost-per-click (CPC) pricing model is main stream in sponsored search, the quality of clicks with respect to conversion rates and hence their values to advertisers may vary considerably from publisher to publisher in a large syndication network. Traffic quality shall be used to establish price discounts for clicks from different publishers. These discounts are intended to maintain incentives for high-quality online traffic and to make it easier for advertisers to maintain long-term bid stability. Conversion signal is noisy as each advertiser defines conversion in their own way. It is also very sparse. Traditional way of overcoming signal sparseness is to allow for longer time in accumulating modeling data. However, due to fast-changing conversion trends, such longer time leads to deterioration of the precision in measuring quality. To allow models to adjust to fast-changing trends with sufficient speed, we had to limit time-window for conversion data collection and make it much shorter than the several weeks window commonly used such as for instance in [3]. Such shorter time makes conversions in the training set extremely sparse. To overcome resulting obstacles, we used two-stage regression similar to hurdle regression [2]. First we employed logistic regression to predict zero conversion outcomes. Next, conditioned on non-zero outcomes, we used random forest regression to predict the value of the quotient of two conversion rates. Two-stage model accounts for the zero inflation due to the sparseness of the conversion signal. The combined model maintains good precision and allows faster reaction to the temporal changes in traffic quality including changes due to certain actions by publishers that may lead to click-price inflation.

Categories and Subject Descriptors

I.5.2 [**Design Methodology**]: Pattern analysis

General Terms

Algorithms, Experimentation

Keywords

Online advertising; machine learning; conversion

1. INTRODUCTION

Large publisher networks represent opportunities for online advertisers to increase their audiences. For example, both Google AdSense and Microsoft AdCenter have large syndication networks

generating billions in revenue each year. However, such syndication networks also present challenges due to significant variation in traffic quality among various publishers comprising such networks. Maintaining separate campaigns for each publisher represents a costly overhead, particularly for smaller advertisers. To address these issues, large online ad-networks introduced tools to adjust pricing of clicks based upon the traffic quality of partner-publishers. Examples are Google's "Smart Pricing" and Yahoo!'s "Quality-based Pricing". Most such smart pricing solutions account for any significant loss of traffic quality from a publisher by giving advertisers discounts based on estimated difference of respective publisher's traffic quality as compared to some benchmark. To measure traffic quality most ad-networks use conversion tracking by advertisers. Conversion is an industry term referring to mechanism where advertisers place beacons in their sites to register when customers, subsequent to an ad-click, perform certain action indicating higher level of engagement with advertiser's product or service. Such action may require actual purchase and then it is frequently referred to as "hard conversion". In other cases, conversion may be registered following an action as common as filling up a form where for instance, customer has provided advertiser with her email address. The latter actions have higher rate of occurrence and are sometimes referred to as "soft conversions". These differences in the method advertisers use to define conversions present a unique challenge for training predictive models. In addition, conversions are very sparse adding to the challenges for developing practical predictive models.

2. OVERVIEW

This work is aimed at finding a predictive model such that given data characterizing particular group of clicks, one would be able to adjust value of these clicks in some economic fashion to the implicit traffic quality expectation expressed through the advertisers' bids. The assumption is that advertisers base their CPC price on certain expectation of conversion likelihood by using their previous history of conversions with more common publishers. Such "standard" publishers are referred here as benchmarks. For most ad-networks their owned and operated sites are frequently used as benchmarks. We describe a two-stage model to predict the adjustment factor. First we estimate the probability for a Query-Publisher-Advertiser (QPA) tuple to have zero conversions. In the second stage, conditioned on at least one conversion, we predict the relative value of a click compared to the click otherwise being the same but originating at the benchmark publisher.

2.1 Dependent Variable

To train a model it is important to find a statistics based on observed quantities that are aligned with the relative traffic quality.

This relative quality can be expressed through a factor F, such that for each tuple in the advertiser and query hierarchies it equalizes the cost of customer acquisition from benchmark publisher with that of a partner publisher with different traffic quality:

$$CPA_j^b = F_j \times CPA_j^p, \qquad (1)$$

where CPA_j^b is the cost of acquisition at benchmark b for query-advertiser pair j, and CPA_j^p is the cost of acquisition at publisher p. Cost CPA_j^b can be expressed as:

$$CPA_j^b = CPC_j \times C_j^b/A_j^b, \qquad (2)$$

where CPC_j is the cost-per-click, C_j^b is the number of clicks and A_j^b is the number of acquisitions for the jth query-advertiser pair. Similarly, the cost CPA_j^p can be written respectively:

$$CPA_j^p = CPC_j \times C_j^p/A_j^p. \qquad (3)$$

Note that CPC_j value is the same in both Eqs. (2) and (3), because advertisers bid the same amount regardless of the publisher that ends up showing their ads. Hence substituting Eqs. (2) and (3) into (1) and simplifying for F_j we get:

$$F_j = \left(C_j^b/A_j^b\right) \times \left(A_j^p/C_j^p\right). \qquad (4)$$

Here we make the assumption that, because each advertiser defines conversion in their own way, the observed conversions CV can be expressed as an advertiser-specific multiplier D_A times the number of actual acquisitions (not observed) as follows:

$$A_j^b \times D_A = CV_j^b, \qquad (5)$$
$$A_j^p \times D_A = CV_j^p. \qquad (6)$$

Using this assumption we can see that F_j can be estimated as:

$$F_j = \left(C_j^b/CV_j^b\right) \times \left(CV_j^p/C_j^p\right). \qquad (7)$$

This observed value is referred hereafter as $CVRI$ (Conversion Rate Index). We are going to use it as dependent variable, conditionally on having at least one conversion.

3. LOGISTIC REGRESSION MODEL

As discussed above we first used logistic regression to model zero conversion outcomes for a QPA tuple using

$$P_j(y = 0|\mathbf{x} = \mathbf{x}_j) = \frac{\exp(\mathbf{x}_j^\top \beta)}{1 + \exp(\mathbf{x}_j^\top \beta)}, \qquad (8)$$

where $P_j(y = 0|\mathbf{x} = \mathbf{x}_j)$ is the probability for the jth QPA tuple to have zero conversions given input vector $\mathbf{x}_j$.

4. RANDOM FOREST MODEL

As discussed above, conditioned on non-zero conversions in a QPA tuple, we set to develop a model to predict the relative value of clicks from that tuple compared with the value of the similar clicks originating from the corresponding benchmark sites. Note that the relative factor represents a continuous real number, in majority cases between 0 and 1, but is allowed to be greater than 1. Hence, we are going to use random forest (RF) regression that minimizes the squared loss R in logarithmic scale:

$$R = \sum_{j:CV_j>0} \left(\log(CVRI_j^{obs}) - \log(CVRI_j^{pred})\right)^2, \qquad (9)$$

where $\log(CVRI_j^{obs})$ is computed for all QPA-tuples with one or more conversions in the following way:

$$CVRI_j^{obs} = \frac{\sum_{m=1}^{C_j} Conv_m/C_j}{\sum_{k=1}^{CB_j} BConv_k/CB_j}, \qquad (10)$$

where C_j is the number of clicks in the jth QPA-tuple, CB_j is the number of clicks from the corresponding benchmark, $Conv_m$ is zero-one function that has value of 1 for clicks that have conversion and 0 for clicks with no conversion; and similarly $BConv_k$ is a signal that has value of 1 for benchmark clicks followed by conversion and 0 for benchmark clicks without conversion.

5. TWO-STAGE REGRESSION

To compute the expected value of predicted relative F_j value for any arbitrary QPA tuple we note that conditional values of F_j for both outcomes predicted by logistic regression model are known.

$$\mathbb{E}[F_j|Conv = 0] = 0, \qquad (11)$$
$$\mathbb{E}[F_j|Conv > 0] = CVRI_j^{pred}, \qquad (12)$$

where $CVRI_j^{pred}$ is based on prediction of the RF model. Since probability P_j and $(1 - P_j)$ are predicted by the logistic regression:

$$\mathbb{E}[F_j] = \mathbb{E}[F_j|CV_j = 0]P_j + \mathbb{E}[F_j|CV_j > 0](1 - P_j). \qquad (13)$$

By substituting Eqs. (11) and (12) into (13) we arrive at:

$$F_j^{pred} = \mathbb{E}[F_j] = CVRI^{pred}(1 - P_j). \qquad (14)$$

Hence by combining predictions of the two models we can compute $\mathbb{E}[F_j]$ for any arbitrary QPA tuple using Eq. (14).

Our experiments showed that the two-stage model was able to achieve 94% variance explained on validation data with mean-squared error (MSE) value of 0.46. We compared this with the current production model based on hierarchical Bayesian updating with 81% of variance explained and with MSE value of 1.25. This improvement was achieved while training data came from a period that was significantly shorter than data used in the baseline models as well as in similar models described in [1] and [3].

6. REMARKS

This work indicates that it is possible to build a practical, large-scale model for dynamic relative quality of the clicks originating from partner-publisher sites in comparison with clicks originating from benchmark sites. This is in contrast with the need for longer history data (several weeks typically), that is required by many traditional statistical models. Another interesting outcome of this effort was to show that the two-stage regression leads to a reasonable precision without compromising performance on tuples with relatively low click volume.

7. REFERENCES

[1] D. Agarwal, R. K. Agrawal, and K. Nagaraj. Estimating rates of rare events with multiple hierarchies through scalable log-linear models. *Proc. of the 16th ACM SIGKDD*, pages 213–222, 2010.

[2] A. C. Cameron and P. K. Trivedi. Hurdle and zero inflated models. *Regression Analysis of Count Data*, pages 125–128, 1998.

[3] R. Rosales, H. Cheng, and E. Manavoglu. Post-click conversion modeling and analysis for non-guaranteed delivery display advertising. *Proc. of the 5th ACM WSDM*, pages 293–302, 2012.