A Joint Optimization of Incrementality and Revenue to Satisfy both Advertiser and Publisher

Dmitry Pechyony, Rosie Jones and Xiaojing Li
MediaMath Inc., 245 1st Street, Cambridge, MA, 02142, USA
{dpechyoni,rjones,xli}@mediamath.com

ABSTRACT

A long-standing goal in advertising is to reduce wasted costs due to advertising to people who are unlikely to buy, as well as to those who would make a purchase whether they saw an ad or not. The ideal audience for the advertiser are those incremental users who would buy if shown an ad, and would not buy, if not shown the ad. On the other hand, for publishers who are paid when the user clicks or buys, revenue may be maximized by showing ads to those users who are most likely to click or purchase. We show analytically and empirically that an optimization towards one metric might result in an inferior performance in the other one. We present a novel algorithm, called `SLC`, that performs a joint optimization towards both advertisers' and publishers' goals and provides superior results in both.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data Mining*

Keywords

digital advertising, multi-objective optimization

1. INTRODUCTION

In display advertising advertisers target users by showing graphical ads on publishers' sites. After seeing an ad the user might perform a desired action (called a *conversion*) on the advertiser's site, for example a purchase. A common ad serving strategy, called *purchase modeling*, is to show an ad if the user has a large probability of purchase. We assume a CPA pricing model, where publisher is paid for conversions. By maximizing the number of conversions, the purchase modeling strategy serves publisher needs.

Surprisingly, purchase modeling might reduce the advertiser's revenue. Suppose that the user is going to purchase in any case. Then, because of advertising costs and possible discounts (e.g. free shipping) that are applied to the purchase that is due to the ad, the advertiser's revenue from this user without showing an ad is larger than the one when the ad is shown. This issue is addressed by another ad serving strategy, called *incrementality modeling*, where the user is shown an ad if this user will convert if shown an ad and will not convert if she is not shown an ad. Such users are called *incremental users* and their conversions are referred to as *incremental conversions*. Thus, with incrementality modeling the advertiser tries to obtain revenue from the users that probably would not bring any revenue were they not shown an ad. For the advertiser, maximizing the number of incremental users is the optimal strategy.

To measure the number of incremental conversions, we use *A/B test*, where the *test users* are shown an advertiser ad and the *control users* are shown a public-service announcement (PSA). We count the number of conversions of users seeing an ad or PSA and estimate the incrementality as:

$$\#\text{ conversions(ads)} - \frac{\#\text{ ads}}{\#\text{ PSAs}} \cdot \#\text{ conversions(PSAs)}. \quad (1)$$

We consider the following commonly used model of interaction between the advertiser and the end users. The advertiser pays to the publisher a fixed amount X for each conversion generated by users who were shown the advertiser's ad by the publisher. Also, the advertiser pays for at most Y conversions. The *advertising budget* is $X \cdot Y$. This is also the maximal revenue that the publisher can earn from the advertiser. Given a fixed budget, the publisher can afford to profitably show ads to the *target list* of top-scored users from its database. A subset of the target list users will be observed online and will actually be shown the ad. *To optimize both incrementality and the publisher's revenue we need to include in the target list both incremental users and users who are likely to buy.*

2. ANALYTIC ANALYSIS

We show that performing just purchase modeling or just incrementality modeling might not be enough in order to maximize both incrementality and the number of conversions. Let D be the set of all users in publisher's database, $P \subset D$ be the set of converters and $I \subset P$ be set of incremental users. Suppose that the publisher can identify both P and I. Let b be the budget. We assume that the size of the target list T is fixed. To measure incrementality we show ads to a random subset of $k|T|$ test users $(0 < k < 1)$ from T. We show a PSA to the rest $(1-k)|T|$ control users. To simplify our analysis, let $|T| = b/k$.[1] We assume that $|I| < b/k < |P|$. The right-hand inequality means that in principle it is possible to spend the entire budget. The left-hand inequality implies that the number of incremental users is not too large.

Consider the purchase modeling approach. Since P is known, this approach will give the same high score to all

[1] Our analysis can be extended to other values of $|T|$ ([1]).

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

users in P and the same low score to all users in $D \setminus P$. We assume that since $|P| > b/k$, a random subset of size b/k from P is in T. Thus the number of conversions due to ads is $\text{Conv}_\text{P} = kb/k = b$. The expected number of incremental conversions is $\text{ConvIncr}_\text{P} = k|I||T|/|P| = b|I|/|P|$.

Now consider the incrementality modeling approach. It gives the same high score to all users in I and the same low score to all users in $D \setminus I$. Since $|I| < |T|$, all incremental users are in T and the number of incremental conversions is $\text{ConvIncr}_\text{I} = k|I|$. Since $|I| < |T|$, a random subset of size $b/k - |I|$ from $D \setminus I$ is in T. Thus the expected number of conversions due to the ads is $\text{Conv}_\text{I} = k(|I| + (b/k - |I|)|P \setminus I|/|D \setminus I|)$. Since $|P \setminus I|/|D \setminus I| < 1$, $\text{Conv}_\text{I} < \text{Conv}_\text{P}$. Also, since $b/k < |P|$, $\text{ConvIncr}_\text{I} > \text{ConvIncr}_\text{P}$. Hence with purchase modeling we maximize the number of conversions but get suboptimal number of incremental conversions. With incrementality modeling we maximize the number of incremental conversions but get a suboptimal number of conversions. We can jointly maximize both the number of conversions and the number of incremental conversions: we put in T incremental users from I and fill up the rest of T with the non-incremental converters from $P \setminus I$. In this case we have b conversions and $k|I|$ incremental conversions.

3. SINGLE LINEAR CLASSIFIER (SLC)

Consider the following reduction [3] from incrementality modeling to binary classification. Let y and y' be respectively conversion and incrementality labels of example x. $y' = 1$ iff x is potentially an incemental example. All examples can be divided into 4 types: i) Potentially incremental examples with $y = 1$ and $y' = 1$, that were shown an ad and converted; ii) Non incremental examples with $y = 0$ and $y' = 0$ that were shown an ad and did not convert; iii) Non incremental examples with $y = 1$ and $y' = 0$ that were shown PSA and converted; iv) Potentially incremental examples with $y = 0$ and $y' = 1$ that were shown PSA and didn't convert. Thus if x was shown an ad then $y' = y$. Otherwise, $y' = -y$. Given a training set $\{x_i, y_i\}_{i=1}^n$ with conversion labels and indicators $\{a_i\}_{i=1}^n$ if the ad was shown, we can build a set $\{x_i, y_i'\}_{i=1}^n$ with the incrementality labels.

We use this reduction to create a novel algorithm, called `SLC`. Let $\text{le}(x) = \log(1 + \exp(x))$. The regularized logistic regression for $\{x_i, y_i'\}$ is $\min_w C \sum_{i=1}^n \text{le}(-y_i' w^T x_i) + \|w\|_2^2$, where $C > 0$ is a hyperparameter. Let r_1 and r_2 are fixed numbers of training examples who were shown an ad/PSA respectively Using the relation between a_i, y_i and y_i' we get a new incrementality modeling algorithm: $\min_w \|w\|_2^2 + C\Big(\sum_{a_i=1} \text{le}(-y_i w^T x_i) + \frac{r_1}{r_2} \sum_{a_i=0} \text{le}(y_i w^T x_i)\Big)$. We replace the fixed ratio r_2/r_1 with a hyperparameter $t > 0$ and get a version of logistic regression that jointly optimizes incrementality and the number of conversions:

$$\min_w C\Big(\sum_{a_i=1} \text{le}(-y_i w^T x_i) + \frac{1}{t} \sum_{a_i=0} \text{le}(y_i w^T x_i))\Big) + \|w\|_2^2. \quad (2)$$

If t is large then the error over PSA examples is ignored and (2) maximizes the number of conversions, thus serving publisher goals. If $t = r_1/r_2$ then (2) does incrementality modeling, thus serving advertiser goals. By tuning t we can optimize both incrementality and the number of conversions. Given a vector w found by (2) and a new example x, the score of x is $w \cdot x$. We denote the algorithm (2) as a single linear classifier (`SLC`). We solved (2) using LIBLINEAR [2].

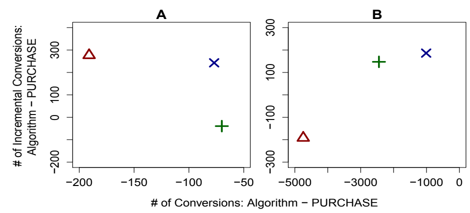

Figure 1: Incrementality/# of conversions of algorithms relatively to `PURCHASE`, × - `SLC`, + - `DIFF-PUR`, △ - `META`. The definitions of the algorithms are in sections 3 and 4.

4. EXPERIMENTS WITH DISPLAY ADS

We did offline experiments with historic data from campaigns of advertisers A and B. In these campaigns ads and PSAs were shown only to the users who visited the advertiser's site in the last 30 days. We defined a conversion as a purchase within 7 days after seeing an ad/PSA. For each campaign we defined 7 non-overlapping days of training and test time periods, where ads/PSAs are shown. Both periods were followed by a 7-day conversion window and were preceded by a 30 day feature window. The training conversion window and test activity window did not overlap. The training/test sets had several million examples. Each user was represented by a single high-dimensional sparse example.

We compared `SLC` with the difference of purchase models algorithm (`DIFF-PUR`) [4]. We used 2-fold cross-validation to tune up the hyper-parameters to maximize both conversion rate and incrementality at the top 50% examples with the highest scores. Additional baselines are original purchase modeling with hyperparameters tuned to optimize the purchase rate (called `PURCHASE`) and a modified purchase modeling where hyperparameters tuned as in `SLC`. We call the latter scheme `META`. All training schemes were based on regularized logistic regression. Fig. 1 shows the results. `PURCHASE` achieved the best number of conversions. But the incrementality of `SLC` is better than that of `PURCHASE`. Also `SLC` has better number of conversions and incrementality than `DIFF-PUR` and `META`. In the full version [1] we show results with two more advertisers, additional schemes for generating training sets and tuning hyperparameters, and provide a theoretical justification for `SLC`.

We have shown that our novel algorithm `SLC` is superior for the important case when we wish to have both large numbers of conversions and good incrementality. In the future we plan to test `SLC` on live traffic. It would be interesting to modify the objective function of `SLC` to optimize directly the incrementality and the number of conversions in top k% of the examples with the highest scores.

5. REFERENCES

[1] http://www.cs.technion.ac.il/∼pechyony/incr_full.pdf.

[2] H.-F.Yu, C.-J.Hsieh, K.-W.Chang, and C.-J.Lin. Large linear classification when data cannot fit in memory. In *KDD*, pages 833–842, 2010.

[3] M.Jaskowski and S.Jaroszewicz. Uplift modeling for clinical trial data. ICML Workshop on Clinical Data Analysis, 2012.

[4] V.Lo. The true lift model - a novel data mining approach to response modeling. *SIGKDD Explorations*, 4(2):78–86, 2002.