Stochastic Models for Budget Optimization in Search-Based Advertising

Paul Ringstad

April 30, 2009

Stochastic Models for Budget Optimization in Search-Based Advertising

- S. Muthukrishnan, Martin Pal, Zoya Svitkina
- WINE 2007

Search-based Advertisements

- Search-based advertising matches users' queries with advertisements
- Advertisers obtain a location near the search results by participating in an auction with other advertisers
- These auctions take place each time a relevant query is received
- Google, for example, receives 10^8 queries per day
- There are many, many small advertisers and advertisers with limited budgets
- This work looks to address expected return on investment for advertisers in this environment of search-based advertising!

Logistics

- Advertiser determines daily budget and the set of relevant keywords
- Relevant keywords and their interaction with users is hard to model
- Likewise, the daily budget is advertiser specific
- Effectiveness of a campaign might include user impressions or even depend on wording of ad
- These complex interactions are hard to model
- Instead, effectiveness of campaign $\equiv$ number of clicks received from users
- We assume all other advertisers' budgets and bids are fixed
- What is the best way to bid on the keywords if we know the cost and expected number of clicks?

Specification

- Advertiser's keywords $i \in T$
- Bids on those keywords, b_i for $i \in T$, can be fractional, $b_i \in [0, 1]$ or integral, $b_i \in \{0, 1\}$.
- b_i is a decision variable for bidding above some threshold amount for keyword i. For example, to bid above the threshold with probability 1 for a keyword i, $b_i = 1$
- $b_i \cdot clicks_i$ is the number of clicks for keyword i where$clicks_i$ is the expected clicks received for winning the auction
- $cpc_i = \frac{clicks_i}{cost_i}$ is the cost-per-click for keyword i
- cpc can be determined for a bidding strategy, b, as in $cpc(b)$

Accommodating a Budget

- Daily budget B
- Instead of limiting strategies to those within the budget, scale the expected clicks by the proportion of the day the strategy will be under budget
- Assume queries and clicks are uniformly distributed throughout the day
- If budget $B < cost(b)$ then we expect $clicks(b) \cdot \frac{B}{cost(b)}$
- Otherwise, we get all the clicks for the keywords that we bid on, $S = \{i | b_i > 0\}$

Budget Optimization

More formally,

- The advertiser seeks to maximize an objective, in this case

$$value(b) = \frac{B}{\sum_{i \in T} b_i cost_i} \sum_{i \in T} b_i clicks_i$$

- This is equivalent to maximizing $clicks(b)$ if $cost(b) \leq B$
- Or, minimizing $cpc(b)$ when $cost(b) > B$

Models

- Evaluate four models which specify the behavior of $clicks_i$
- One Fixed model, where $clicks_i$ is deterministic, and three Stochastic models for which $clicks_i$ is stochastic
- Evaluation: Given a bidding solution, can we evaluate the expected value?
- Optimization: Can we determine a bidding solution that maximizes the expected value of an objective function?
- This is trivial for the Fixed model, but not for the others
- For each of the models we will address these two concerns

The Fixed model

- $clicks_i$ is a predetermined integer
- Evaluation: is trivial given a bidding solution b

$$value(b) = \frac{B}{\sum_{i \in T} b_i cost_i} \sum_{i \in T} b_i clicks_i$$

- Optimization: By reduction to the fractional knapsack problem the optimal solution is a prefix solution

Prefix Solutions

- Given a set of keywords ordered by increasing cpc_i
- For an *integer* prefix solution there exists some i^* s.t. $b_i = 1$ for all $i \leq i^*$ and $b_i = 0$ for all $i > i^*$
- For a *fractional* prefix solution there exists some i^* s.t. $b_i = 1$ for all $i < i^*$ and $b_i = 0$ for all $i > i^*$, $b_{i^*} = [0, 1]$
- Mandates bidding on the "cheap" keywords

Prefix Solution for Fixed Model

- Find the maximum index i^* s.t.

$$\sum_{i \leq i^*} cost_i \leq B$$

- If i is the last index in T, set $b_i = 1$ for all i
- Otherwise, find $\alpha = [0, 1)$ s.t.

$$\sum_{i \leq i^*} cost_i + \alpha cost_{i^*+1} = B$$

The Fixed model

Theorem: The optimal fractional solution for the BO problem in the Fixed model is the maximal prefix whose cost does not exceed the budget.

Proof.

Given a solution $b' \neq b$, if $cost(b') < B$ we can improve the value by slightly increasing our bid on a keyword as that will increase the number of clicks. If $cost(b') > B$ we can improve the value by decreasing the bid on the most expensive keyword, lowering the effective $cpc(b)$. □

Summary of the Fixed model

- Fractional prefix solution is optimal
- By reduction to integer knapsack problem, integer prefix solution is NP-hard

The Proportional Model

- Given q_i, the proportion of total clicks for a keyword i
- Random variable C determines the total number of clicks
- $clicks_i = q_i C$
- In the stochastic setting, the objective function becomes

$$E[value(b)] = E[\frac{1}{max(1, cost(b)/B)} \sum_{i \in T} b_i clicks_i]$$

Evaluating the Proportional Model

- The solution depends on the value of C
- Define a threshold

$$c^* = \frac{B}{\sum_{i \in T} b_i q_i cpc_i}$$

- s.t. for $c \leq c^*$, $cost^c(b) \leq B$ and for $c > c^*$, $cost^c(b) > B$
- The objective function becomes

$$E[value(b)] = \sum_{i \in T} b_i q_i \sum_{c \leq c^*} cp(c) + \frac{B}{cpc(b)} Pr[C > c^*]$$

- assuming $Pr[C > c^*]$ and $\sum_{c \leq c^*} cp(c)$ can be easily evaluated

Prefix Solution is an Optimal Solution

Theorem: The optimal fractional solution for the SBO problem in the Proportional model is a fractional prefix solution

Proof.

Given a solution b, if b is not a prefix solution then there is some keywords $i < j$, $b_i < 1$, $b_j > 0$. Choose the smallest such i and the largest such j. If $q_i cpc_i = 0$ set $b_i = 1$ and continue. Otherwise, pick the maximum δ_i,$\delta_j > 0$ that satisfy $\delta_i \leq 1 - b_i$, $\delta_j \leq b_j$, $\delta_i = \frac{q_j cpc_j}{q_i cpc_i}\delta_j$. If we set $b'_i = b_i + \delta_i$ and $b'_j = b_j + \delta_j$ and $b'_k = b_k$ where $k \notin \{i, j\}$, then solution b' has the property $cost^c(b') = cost^c(b)$ and $clicks^c(b') \geq clicks^c(b)$.

□

Optimizing in the Proportional Model

Theorem: The optimal fractional solution for the SBO problem in the Proportional model can be found exactly in time O(n+t), where t is the number of possible values of C, or approximately by a PTAS.

- The proof consists of "marking" points in the space of possible prefixes, for example all integer prefixes and all threshold prefixes that spend the exact budget.
- This partitions space of prefixes into intervals. If b and b' are in the same interval I, the set of C values that will cause those solutions to be over budget is the same.
- Added to the "marked" points, "interesting" points consist of optimal prefix solutions *within* these intervals
- To find the optimal prefix, all "marked" and "interesting" points are evaluated

Summary of the Proportional model

- Fractional prefix solution is optimal
- Exact optimal solution can be found in $O(n+t)$ for polynomial-sized C
- PTAS exists for continuous C

The Independent Model

- Each keyword $i \in T$ has a probability distribution over number of clicks, p_i.
- The variables $clicks_i$ and $clicks_j$ are independent for all $i \neq j$
- Realistic for keywords belonging to different topics, i.e. flowers and sports

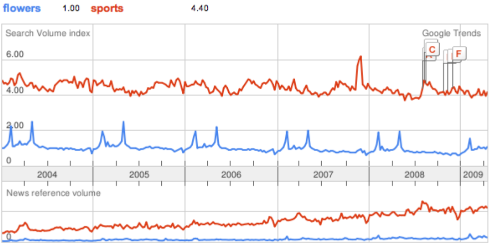

Figure: Comparison of query traffic for keywords "flowers" and "sports" on Google Trends

Evaluating a Solution in the Independent Model

$$E[value(b)] = \sum_{i \in T} \sum_{c \in C_i} p_i(c) b_i c \sum_{d \geq 0} \frac{1}{f_i(c,d)} Pr[cost(b_{-i}) = d]$$

- where $cost(b_{-i}) = \sum_{j \neq i} b_j cpc_j clicks_j$
- $f_i(c,d) = max(1, \frac{d + c \cdot cpc_i}{B})$

Lemma: For any given $\epsilon > 0$, there is a polynomial-time algorithm that finds a value s' such that $s(i,c) \leq s' \leq (1+\epsilon)s(i,c)$ where

$$s(i,c) = \sum_{d \geq 0} \frac{1}{f_i(c,d)} Pr[cost(b_{-i}) = d]$$

This is achieved by setting the range of possible values of d from 1 to $M = \sum_{i \in T} max\{c \cdot cpc_i | c \in C_i\}$, with intervals of size $(1 + \frac{\epsilon}{n})^k$

The Independent Model and Fractional Prefix Solution

Theorem: In the Independent model of the SBO problem, the optimal fractional solution may not be a prefix solution

Proof.

Given three keywords, let $cpc_1 = 0$, $cpc_2 = 1$, $cpc_3 = 1$, $clicks_1 = 1$ with probability 1, $clicks_2 = (0 or 1)$ with probability $\frac{1}{2}$ and $clicks_3 = 1$ with probability 1. The optimal solution is $b_1 = b_3 = 1$ and $b_2 = 0$ which yields 2 clicks. However, the best prefix solution is $b_1 = b_2 = 1$ and $b_3 = 0$ which gets 1 or 2 clicks, each with probability $\frac{1}{2}$ □

Integer Prefix Solution for Independent Model

- *Theorem:* For any integer solution b to the SBO problem in the Independent model, there exists an integer prefix solution b_v s.t. $E[value(b_v)] \geq \frac{1}{2}E[value(b)]$. In particular, the solution b_v bidding on the set $V = \{i | i \leq i^*(b)\}$ has this property
- Proven using truncated solution b_U that preserves this property and then "filling in the gaps" in b_U to form a solution b_V that also maintains this property
- Using the above result, and the $(1+\epsilon)$-approximation,
- *Theorem:* There is a $(2+\epsilon)$-approximation algorithm for the SBO problem in the Independent model, which runs in time polynomial in n, $\frac{1}{\epsilon}$ and $logM$
- where M is the maximum cost for all the clicks for keyword i

The Scenario Model

- A set of scenarios Σ with a probability distribution defined over it
- The number of $clicks_i$ for $i \in T$ are determined by different *scenarios*, σ where $\sum_{\sigma \in \Sigma} p(\sigma) = 1$

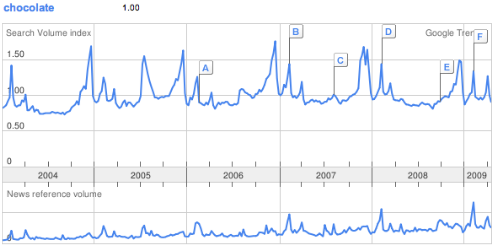

Figure: Query traffic for keyword "chocolate" on Google Trends

Negative Results for the Scenario Model

- *Theorem:* The SBO problem is NP-hard in the scenario model
- By reduction from Clique, both integer and fractional solutions are NP-hard
- *Theorem:* The gap between the optimal fractional prefix solution and the optimal (integer or fractional) solution to the SBO problem in the Scenario model can be arbitrarily large.
- This can be shown by setting up the scenarios s.t. even-numbered keywords are c-times more expensive than their preceding odd-numbered keywords. The optimal solution for SBO involves bidding only on odd keywords and the prefix solution can be arbitrarily far from optimal as c and n increase.

Extensions

- Generalizes to a model which takes into account probability of a user purchasing after clicking for a given keyword
- Generalizes to the k-slot advertising environment as follows:
- For Fixed and Proportional models prefix solution is still optimal
- Independent model breaks since keywords become "landscapes", where $j \in k$ slots results in different cpc_i^j values for a keyword i
- Scenario model is still NP-Hard

Wrap-up and Discussion

- Are advertisers satisfied with a prefix campaign?
- Is there a way to incorporate maximizing coverage into the objective function?
- Other, more expressive models?
- Methods for approaching the scenario model
- Methods for optimization through learning: adaptive, budget constrained algorithms
- Budgeted Multi-Arm Bandits seems to contain some of the right elements of such an algorithm