Methodology and Challenges in Search Marketing at Business.com

Nathan Janos, Tom Anderson, Paul Dagum
Business.com

Abstract

Annual search marketing spending in the U.S. grew 28% to US$7.9B in 2006, outpaced traditional marketing and is expected to double by 2010. The large return on investment has led to an increase in demand for search terms while posing new challenges in running a profitable campaign. Campaigns must systematically and repeatedly analyze large amounts of click data, adjust bid inferences across thousands of search terms, maximize conversion statistics and best monetize ad-dollar spend.

At Business.com we have developed an econometric model for search marketing. We model demand elasticity using generalized linear and mixed models with random cluster effects across groups of correlated search terms and repeated measures across search engine partners. Bid positions are adjusted daily using either a profit optimization or a portfolio revenue optimization across more than 100,000 search terms.

We discuss these techniques, information challenges in the production platform and our results in managing a successful search marketing campaign.

KEY WORDS: search marketing, econometrics, generalized linear model, linear mixed model, generalized linear mixed model

1. Introduction

As a search engine and directory for the business-to-business (B2B) search vertical, Business.com desires to increase the volume of visitors to its pages and listings. The two primary methods for acquiring traffic are through natural (organic) search listing results and through paid advertisement placement. This paper concerns itself with the latter form of traffic acquisition: paid placement, also known as search engine marketing (SEM).

1.1 The Business-to-Business Vertical Market

Business.com is an online directory and search engine of business clients that wish to direct their advertisements to the people who have purchasing power at small-, medium- and large-sized businesses. It is a directory of business solutions for people who run businesses. Within this B2B market, also known as a market vertical, Business.com is one of the largest directories of listings and clients. Business.com has made its mark in the B2B advertising market through consistently offering highly qualified traffic to its clients. By acting as a search engine and qualifying traffic through natural and paid traffic funnels Business.com is able to effectively filter traffic and greatly increase return on investment (ROI) not only from Business.com's paid purchasing standpoint, but also from Business.com's clients' standpoints.

1.2 Motivation

In the first quarter of 2006 SEM sources of traffic accounted for 45% of Business.com visits and 56% of site revenue. A third party formerly managed a large portion of Business.com's paid advertisement placement. The approach described in this paper was developed in order to avoid paying third-party agency fees that were as high as 10% of advertising spend. Business.com accurately identified the strategic value of developing core compentency in search marketing. This development process involved a combination of talent, people, processes and technologies. Also, importantly, Business.com wanted to respond more quickly to internal requests from sales and marketing and to have full control over testing procedures that involve SEM bidding.

The problem of developing an SEM purchasing system posed a number of computational and statistical challenges. These challenges can be decomposed into two broad areas: the management of large amounts of data and the application of an efficient purchasing model to this data. This paper discusses the econometric model used and also the production system deployment of this model.

1.3 Other Approaches

A variety of companies have developed SEM purchasing systems and services. These companies, such as MatchCraft, Adapt and Efficient Frontier, offer bid management, ROI analysis and other proprietary profit and revenue optimization services. These systems do not always offer individualized approaches or the desired flexibility. It may be in the best interest for larger companies seeking out SEM optimization to develop and own internal solutions and the associated intellectual property.

2. Search Engine Marketing

Search engine marketing is a type of direct marketing that focuses on online advertisement distribution instead of more traditional distributions such as television, magazines, newspapers and billboards. SEM is the field of marketing on the internet through search engine portal result screens and also through contextually distributed advertisements in online content. In the first of these cases, a user of a search engine is presented with advertisements (usually text only) next to search results. These advertisements are contextually relevant to their search. In the second case, a user consuming content (consisting of text or other media) is presented with contextually relevant advertising next to the content they are consuming. Both of these methods of advertisement distribution share similarities in that a search algorithm is used to match contextually relevant ads to user behavior.

2.1 Online Advertisement Model

Many different techniques have been applied to online advertisement since the beginnings of the browseable web in the 1990s. Over time more sophisticated techniques have supplanted initial advertisement efforts. The four main components of any advertising model involve the advertisement (usually text or a graphic), the advertiser (or client), the person targeted by the advertisement (a person browsing the web) and finally a third party (such as a search engine) that facilitates the distribution of the advertiser's advertisement to users. Within this framework and in increasing orders of sophistication the primary models are:

- Cost Per Impression (CPM) - The advertiser pays a third party for every 1000 advertisements displayed on users' screens

- Cost Per Click (CPC or PPC) - The advertiser pays a third party for every click by a user on their advertisement

- Cost Per Acquisition (CPA) - The advertiser pays the third party only when a user fulfills a given action as defined by the advertiser, which may include purchasing a product, registering a user or obtaining a lead

One of the benefits of the PPC model is that advertisers can track every click and the price at which every click was purchased. In addition to this, advertisers are able to then track whether a paid click converts into a sale or lead. Some PPC companies, such as Google, even offer conversion tracking for their advertisers and many large companies maintain internal conversion tracking systems. It is in this ability to track every user that PPC marketing distinctly diverges from traditional marketing methods. *With PPC advertising it is possible to accurately report the return on marketing investment.*

The specific form of search engine marketing addressed in this paper is the PPC (or CPC) model. In this model the search engine is paid per user click. We evaluate our own cost of acquisition as well as expected revenue per click and factor this into the price at which paid advertising is purchased on a per-advertisement basis. In effect, this internal tracking of ROI can be considered to be a type of CPA marketing.

2.2 Bidding

The major search engines (including Business.com, Google, Yahoo and MSN) now use a performance-based system for ad placement. Performance-based placement incorporates proprietary methods for evaluating not only the price (bid) that an advertiser is willing to pay but also other performance metrics such as the estimated amount of traffic that might be expected with an advertiser's ad copy. Thus, the driver of real value in the PPC advertising model is the rank on a search engine results page. Drivers that control rank in a performance based model include cost-per-click, click-through-rate and landing page quality. Factors that determine cost-per-click include ad copy, ad relevance and market competition.

Figure 1 shows a screen shot of a Google search result for the keyword 'employee evaluation'. In these results you can see three PPC advertisements listed in a tan box at the top of the search results page and another column of PPC advertisements listed down the right hand side of the screen. Business.com's advertisement can be seen in the top position of the tan box. In the lower left hand side of the screen are organic (or non-paid) search results.

2.3 The Metrics

The following list defines and relates the common metrics in SEM:

- Impression - An impression is counted when an ad is displayed in the user's web browser window

- Click - The event marked by a user single clicking on an advertisement

- Click-through-rate (CTR) - The number of clicks divided by the number of impressions; $CTR = clicks/impressions$

- Cost per click (CPC) - The cost per click, which usually varies from the bid by being less than or equal to the bid; $CPC = cost/click$ and $CPC \leq bid$

- Conversion - The monetization or generation of user action by the marketer

- Conversion rate (CR) - The rate as which visits to a user's web site are converted into monetizeable events; $CR = events/clicks$

- Revenue per visit (RPV or RPI) - The revenue generated by the marketer per user click; $RPV = revenue/click$

- Bid (also Max CPC) - The maximum amount the marketer is willing to pay for an advertisement placement

- Position - The spatial placement of an advertisement on a search results page

2.4 The Keyword

The keyword used to match searches to an advertisement is the smallest measured segment of data. All metrics are calculated for each keyword ad copy on a daily basis. Each of the following metrics is collected across all search engines for each keyword and for each day:

- Ad position for that keyword and ad copy

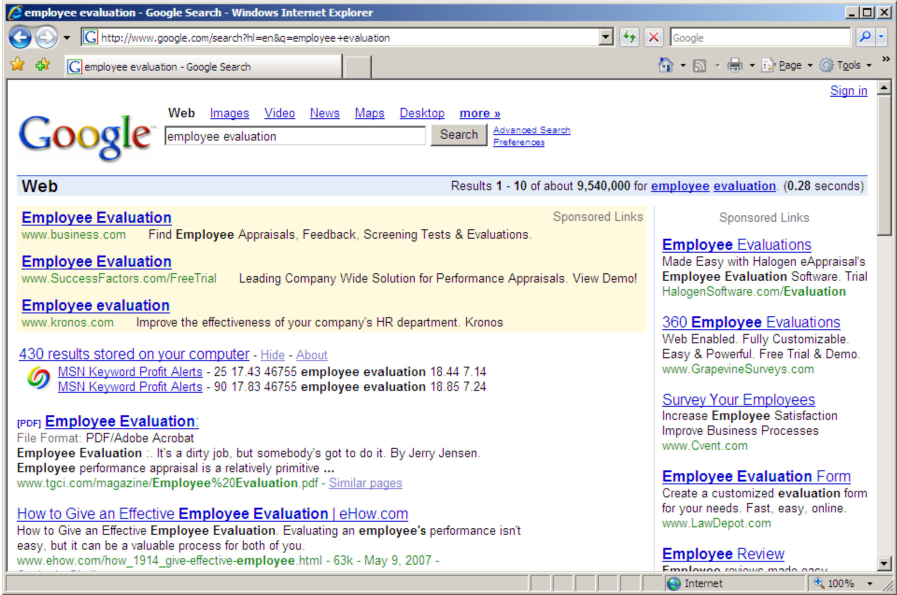

Figure 1: Sample Search Results Page

- CTR on search engine
- CPC on search engine
- Impressions on search engine
- Bid
- Revenue per search engine click from client listing clicks

2.5 SEM Control Factors

SEM optimization occurs through observing and modeling the cost and revenue of each keyword. The cost and revenue can be expressed in terms of the keyword metrics and these keyword metrics are, in turn, influenced by several key factors including:

- Ad copy relevancy and search engine position which affect probability of a search engine click
- CPC is determined by our bid and the associated competitive landscape
- Relevance and page design affect the conversion probability on a Business.com page
- Revenue per Business.com click is affected by advertiser coverage and CPC as well as yield management of client listing inventory
- Time of day, day of week and seasonality which all affect these factors

As mentioned earlier, it is important to note that the keyword bid is the primary quantifiable controller. Thus, controlling and effectively modeling the keyword metrics as a function of bid is the goal.

2.6 Search Marketing Econometrics

The econometrics of search marketing can be expressed in terms of the keyword metrics. Since we know that position is the primary quantifiable controller we will want to express all keyword metrics in terms of position.

The model can be further decomposed from the definition of profit, where profit equals revenue minus cost:

$$\pi = R - C.$$

where

- π is profit.
- R is revenue.
- C is cost.

Revenue and cost can be decomposed into volume times revenue per volume (also called revenue per page view, revenue per impression or RPI) and cost per volume (which, in this case, is the cost per click), where volume is another term for the number of valid clicks received:

$$R = RPI \times VOL$$

$$C = CPC \times VOL$$

$$\pi = VOL \times (RPI - CPC)$$

From this point volume can be decomposed into total number of impressions times the clicks per impression (also called the click-through-rate):

$$CTR = clicks/impressions$$

$$VOL = impressions \times CTR$$

This equation allows us to break down the problem of modeling profit into its primary components, which are the click through rate, impressions, revenue per visitor and cost per click. Each of these components can be considered as a function of bid position, x, since bid position is the primary quantifiable controller in the PPC model:

$$\pi(x) = impressions(x) \times CTR(x) \times [RPI(x) - CPC(x)]$$

These four components—$impressions(x)$, $CTR(x)$, $RPI(x)$ and $CPC(x)$—identify the four primary equations which must be modeled to effectively solve the profit-per-position problem.

3. Demand Chain Model for Online Leads

A demand chain is composed of the enterprises that sell a business's goods or services. For example, a demand chain may be composed of buyers who initiate the sales transaction, the resellers who sell the manufacturer's goods, and the manufacturer who creates the goods. In the case of SEM, an advertisement is purchased per user click and then sold to a client for profit. The probability of selling a click to a client must be included in this demand chain model as well as the probability of acquiring that click. In this scenario, Business.com is both a buyer and a seller of advertising volume on a per click basis.

Figure 2 shows the demand chain model with Business.com occupying the central gathering point of traffic before client distribution. In the figure, cost is generated when traffic flows into Business.com. Revenue is generated when that traffic funnels out to Business.com's clients. Each hop in the graph represents a further qualification of a user lead, which makes it more valuable to the client.

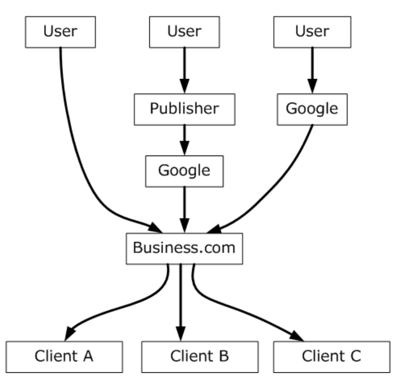

Figure 2: Demand Chain Funnel

3.1 Generalized Linear Model

In the generalized linear model (GLM) described here, the mean of the response variable is modeled as a monotonic nonlinear transformation of a linear function of the predictors. GLMs were first described by Nelder and Wedderburn and then later by McCullagh and Nelder. The inverse of this transformation function is called the link function. The response variable may have a distribution other than a normal distribution and certain distributions lend themselves to particular link functions. The defining elements of a GLM are:

- Predictors
- Response distribution
- Fixed or random effects

Table 1 summarizes the distribution and link function used to describe each GLM-modeled metric.

Metric	Distribution	Link Function
CTR	binomial	logit
CPC	Poisson	log
Impressions	Poisson	log
Bid	Poisson	log

Table 1: Distribution and Link Functions

In addition to the distributions and link functions assumed for these metric models, classifiers and data weighting are also available for use. A configurable parameter can classify more recent data differently and a configurable weighting vector can be supplied to the GLM model to weight more recent data.

3.2 Fixed Versus Random Effects

The decision to treat effects as either fixed or random can be complex in practice. While deriving the search marketing model we identified a number of unobserved hierarchical effects related to the keyword's business category and unobserved individual keyword effects such as ad copy. The random-effects formulation assumes the following:

- Unobserved effects are random draws from a common population
- Explanatory variables are strictly exogenous

Our objective is to infer a user's response to changes in listing position for a keyword and to predict that response for that keyword. Our sample of keywords is the entire universe that we study. Prediction is improved by treating these effects as fixed.

The second assumption of random-effects formulation is also violated in search marketing. We have previously noted that top-tier search engines, including Business.com, use performance ranking of listings that incorporate CTR adjusted for position and bid. The CTR is a function of unobserved effects such as ad copy, the number of competing listings on the page, and the ad copy of these competitors. The bid explanatory variable cannot be strictly exogenous because an improvement in ad copy translates to an improvement in CTR and consequently a decrease in the bid.

Based on these findings we treat the unobserved effects as fixed effects. This simplifies the challenging computational inference by allowing us to use a GLM formulation in lieu of a generalized linear mixed model (GLMM) that requires numerical integration of the random effects.

Effects were chosen in order to leverage both hierarchical structures within the data classification as well as longitudinal time factors such as week-to-week variability. The organization of the keyword data into keywords contained in ad groups (where an ad group defines a particular segment of site traffic) lends themselves to an approach that utilizes both this hierarchical and longitudinal structure of the available data.

A GLM that expresses this ordering can be written as:

$$g(y_{kt}) = \gamma_g + \delta_g x_{kt} + \omega_s + \tau_{h(t)}$$

where

- $g(\dots)$ is the monotonic nonlinear link function associated with the chosen distribution
- y_{kt} is the response for keyword $k = 1, \dots, K$ at time $t = 1, \dots, T$
- x_{kt} is the fixed-effect regressor for keyword k at time t
- γ_g and δ_g are the hierarchical, fixed-effect coefficients for each ad group $g = 1, \dots, G$
- ω_k is the hierarchical, fixed-effect coefficient for a keyword k
- $\tau_{h(t)}$ is the longitudinal, fixed-effect coefficient for a keyword at $h(t)$ weeks in the past
- $h(t) = 1, \dots, W$ indexed weeks in the past

In matrix form and with a slope coefficient for ad groups only this can be expressed as

$$\mathbf{D} = \left[\begin{array}{ccc|ccc|ccc}
X_{11} & 0 & \dots & Z_{11} & 0 & \dots & Z_{11} & 0 & \dots \\
X_{12} & 0 & \dots & Z_{12} & 0 & \dots & 0 & Z_{12} & \dots \\
X_{13} & 0 & \dots & Z_{13} & 0 & \dots & 0 & 0 & \dots \\
\vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \dots \\
X_{21} & 0 & \dots & 0 & Z_{21} & \dots & Z_{21} & 0 & \dots \\
X_{22} & 0 & \dots & 0 & Z_{22} & \dots & 0 & Z_{22} & \dots \\
X_{23} & 0 & \dots & 0 & Z_{23} & \dots & 0 & 0 & \dots \\
\vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \dots \\
0 & X_{31} & \dots & 0 & 0 & \dots & Z_{31} & 0 & \dots \\
0 & X_{32} & \dots & 0 & 0 & \dots & 0 & Z_{32} & \dots \\
0 & X_{33} & \dots & 0 & 0 & \dots & 0 & 0 & \dots \\
\vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \dots \\
0 & X_{K1} & \dots & 0 & 0 & \dots & Z_{K1} & 0 & \dots \\
0 & X_{K2} & \dots & 0 & 0 & \dots & 0 & Z_{K2} & \dots \\
0 & X_{K3} & \dots & 0 & 0 & \dots & 0 & 0 & \dots \\
\vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \dots \\
0 & X_{KT} & \dots & 0 & 0 & \dots & 0 & 0 & \dots
\end{array}\right]$$

or

$$\mathbf{D} = \left[\begin{array}{ccc|ccc|ccc} \mathbf{g_1} & \dots & \mathbf{g_G} & \mathbf{k_1} & \dots & \mathbf{k_K} & \mathbf{w_1} & \dots & \mathbf{w_W} \end{array}\right]$$

and

$$\mathbf{\Theta} = \begin{bmatrix} \Theta_1 \\ \Theta_2 \\ \Theta_3 \\ \vdots \\ \Theta_{2 \times G+K+W} \end{bmatrix}$$

where $g(\mathbf{y})$ can be re-written in terms of $\mathbf{D}$ and $\mathbf{\Theta}$ as

$$g(\mathbf{y}) = \mathbf{D}\mathbf{\Theta}$$

and where

- X_{kt} represents a sub-matrix of size 1×2 of the form $X_{kt} = \begin{bmatrix} 1 & x_{kt} \end{bmatrix}$
- $Z_{kt} = 1$
- Θ_z represents a sub-matrix of size 2×1 where $z = 1, \dots, 2 \times G + K + W$ of the form $\Theta_z = \begin{bmatrix} \gamma_g \\ \delta_g \end{bmatrix}$ for coefficients corresponding to ad groups, $\Theta_z = \omega_k$ for coefficients corresponding to keywords, $\Theta_z = \tau_{h(t)}$ for coefficients corresponding to week categorization.
- $\mathbf{g_g}$ is a column vector of length $K \times T$ which represents the labeling of a regressor and corresponding response variable to a given ad group, g
- $\mathbf{k_k}$ is a column vector of length $K \times T$ which represents the labeling of a regressor and corresponding response variable to a given keyword, k
- $\mathbf{w_w}$ is a column vector of length $K \times T$ which represents the labeling of a regressor and corresponding response variable to a particular week, w
- $\mathbf{\Theta}$ is a column vector of length $2 \times G + K + W$ containing all intercepts and slope coefficients, alternating respectively
- $\mathbf{D}$ is a matrix with $K \times T$ rows and $2 \times G + K + W$ columns which encapsulates input regressors with the inherent hierarchical and longitudinal structures identified and described

An example of a design matrix which encapsulates two adgroups, four keywords, two weeks of data and two data points per keyword (one during each week) may be written as

$$\left[\begin{array}{cccc|cccc|cc}
1 & x_{11} & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & x_{12} & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & x_{21} & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & x_{22} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & x_{31} & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & x_{32} & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & x_{41} & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & x_{42} & 0 & 0 & 0 & 1 & 0 & 1
\end{array}\right]$$

and

$$\Theta = \begin{bmatrix} \gamma_1 \\ \delta_1 \\ \gamma_2 \\ \delta_2 \\ \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \\ \tau_1 \\ \tau_2 \end{bmatrix}$$

where both $\mathbf{D}$ and Θ can be algorithmically generated based on number of ad groups, keywords, days of data, and from how many weeks back the data has been sourced. When variable days of data exist for each keyword the number of rows will vary.

With the problem expressed as $g(\mathbf{y}) = \mathbf{D}\Theta$ a link function and appropriate distribution can be applied and the problem is reduced to one in which the coefficients can be solved for with a standard GLM algorithm.

An example of a design matrix, $\mathbf{D}$, created from real data can be seen in Figure 3 for one campaign, five adgroups, nine keywords and various weekend/weekday and weeks-back periods of time. Figure 3 shows the existence of an X_{kt} or Z_{kt} element in the design matrix as a dot.

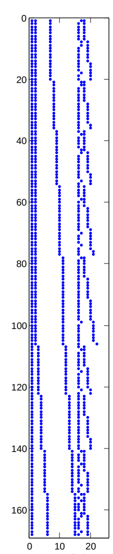

Figure 3: Sample Design Matrix

3.3 Link Functions

The link functions for this problem were specifically chosen to address underlying phenomenon in CTR, CPC, bid and impressions. CTR is modeled with a logit link function and binomial distribution

$$g(\mathbf{y}) = \log\left(\frac{\mathbf{y}}{1-\mathbf{y}}\right) = \mathbf{D}\Theta$$

and the CPC, bid and impression are modeled with a log link function and Poisson distribution

$$g(\mathbf{y}) = \log(\mathbf{y}) = \mathbf{D}\Theta$$

where the design matrix expresses the hierarchical and time-based structures chosen.

3.4 Stochastic Modeling

The SEM PPC marketplace is inherently variable. A model which thrives in this environment is best suited for SEM PPC modeling. Because the market for each keyword changes from week to week, the optimal profit position also changes. Thus, the bidding process is a continual one in which the target is constantly moving.

The models used require data sampling from different segments of the position space. This moving target phenomenon inherent in SEM marketing is thus a good candidate for models which improve with market variability.

3.5 Clustering Sparse Data

When keyword data is too sparse, it can be clustered with other similar keywords. One approach is to take the metrics for sparse keywords that are all directing advertising traffic to a single page. The assumption is that because these keywords are grouped by topic that they might also share other characteristics. The GLM described also allows us to leverage trends across keywords with marginally sparse data.

3.6 Profit Maximization

Profit maximization can be achieved by modeling the profit curve, $\pi(x)$, for each keyword in an account and searching over the modeled profit function for the maximum profit point. Then, by using the modeled difference between CPC and bid, it is possible to find the bid that should correspond to this maximum profit point.

4. The Production System

The production system first gathers cost data (search engine statistics) and revenue data (advertiser statistics) from the appropriate databases. The cost and revenue data are combined and bids are calculated with the appropriate models for each keyword. As the scheduling dictates, these bids are then published back into the database where they are picked up by a web application and uploaded to the search engines.

The production system consists of these three primary processes:

- Importing cost and revenue data from SEM database
- Calculating bids using MATLAB programs
- Exporting bids back to SEM database as scheduled

4.1 Inherent Bidding Limitations

Search engines do not immediately respond to bid changes but rather sample the bidding space with a new bid. Due to this sampling there is usually an inherent lag time between a stable response to a bid change and when the bid changes. This lag time is roughly one to two days. Thus, this limitation means that intraday bidding could be counterproductive and, in fact, it is possible that it can lead to undesired feedback or out-of-phase bidding.

4.2 Problem Complexity

The GLM methods described in this paper can require significant hardware requirements for the analysis of just a nominal number of keywords. In our case we use machines with 64-bit architectures in order to complete matrix manipulations on the design matrix described, $\mathbf{D}$, which has $2 \times G + K + W$ columns and $K \times T$ rows. In the case where:

- $G \approx 2000$
- $K \approx 10000$
- $W \approx 8$
- $T \approx 56$

the design matrix in a case like this can contain over one billion elements. Because $\mathbf{D}$ can be considered as a sparse matrix, this fact can be leveraged in algorithmic generation of $\mathbf{D}$ and in matrix compression routines.

4.3 Data Collection

Data is collected in varying lengths of sequential seven-day segments. In general the number of days of data collected for each keyword is $7n$, where n is commonly 4 to 8. The reason for collecting data in this manner is to normalize day-of-week effects on data such that an equal number of days of data for each day of the week is collected. It is important to recognize that metrics behavior differently on week days and weekends.

4.4 Rules-Based Bidding

Many keywords experience a low volume of click traffic. In these cases it is especially difficult or even impossible to use the GLM models described with any level of acceptable confidence. For these keywords a näive rules-based system has been applied. In these cases the rules-based bidding takes into account several pivot points along the bidding spectrum. These include:

- The RPI or page monetization, pm, ability of a keyword, which is the upper limit on the amount of revenue that can be expected per search engine click and thus the initial upper limit on the bid
- When the RPI data is scarce for a keyword the group RPI for all keywords associated with an adgroup is used
- The minimum activation bid, ma, required by the search engine, which is the lower limit for the bid
- An exploration factor, β, which describes how quickly the bid space between pm and ma is evaluated; $\beta \geq 1$
- An exploration factor, α, which describes how quickly the bid space below the current bid is evaluated; $\beta \leq 1$

The rules are simple and are designed to explore the bid space efficiently while limiting financial exposure. The rules can be described as the following:

- If $\pi \geq 0$ then $bid_{new} = max(bid_{old} \times \beta, ma)$
- If $\pi < 0$ then $bid_{new} = max(bid_{old} \times \alpha, ma)$

This rule-based approach allows for financially safe exploration of the bidding space. In the event that a low-volume keyword is profitable in this space at a higher bid and in the case that this provides sufficient amounts of data the GLM-based model then takes over the bidding process resulting in a successful search by the rules-based system.

5. Results

The GLM formulation described here offers various configurations depending on the ad group segments and time segmentation chosen. Our production system implementation enables us to adapt to market changes over time by applying different configurations to different segments of data as determined to be optimal by our fit and prediction test suites.

Our ad groups (which can also be considered as traffic segments) are divided into high-, medium- and low-volume segments. For each of these groups one of the following GLM model formulations is chosen:

- M1 - Individual slope and intercept
- M2 - Group slope, individual intercept for each keyword
- M3 - Group slope, individual intercept, intercept offset for week segmentation
- M4 - Group slope, individual intercept, intercept offset for weekend/weekday segmentation

Furthermore, Akaike's information criterion (AIC) is used to control for over-parameterization and to help pick a model that is optimally robust in a production environment. In terms of the likelihood function, L, and the number of parameters in a model, k, the AIC can be expressed as:

Model	Mean Fit Resid.	Std. Fit Resid.	Mean Pred. Resid.	Std. Pred. Resid.	AIC
M1	-7.61×10^{-4}	9.78×10^{-3}	-2.47×10^{-3}	1.25×10^{-2}	2391.9
M2	-8.17×10^{-4}	9.83×10^{-3}	-2.69×10^{-3}	1.34×10^{-2}	1261.6
M3	3.79×10^{-4}	9.83×10^{-3}	-1.48×10^{-3}	1.32×10^{-2}	1263.3
M4	7.33×10^{-4}	9.83×10^{-3}	-2.60×10^{-3}	1.34×10^{-2}	1263.5

Table 2: CTR Residuals

Model	Mean Fit Resid.	Std. Fit Resid.	Mean Pred. Resid.	Std. Pred. Resid.	AIC
M1	-2.10×10^{-3}	6.80×10^{-2}	1.73×10^{-2}	9.62×10^{-2}	2561.0
M2	-2.44×10^{-3}	8.07×10^{-2}	2.21×10^{-2}	1.05×10^{-1}	1520.0
M3	-2.86×10^{-2}	8.29×10^{-2}	-4.16×10^{-3}	1.06×10^{-1}	1516.7
M4	-4.63×10^{-3}	8.08×10^{-2}	1.99×10^{-3}	1.05×10^{-1}	1521.3

Table 3: CPC Residuals

$$AIC = 2k - 2\log L$$

The deviance of the fit, D, can be expressed as:

$$D = -2\log L$$

Allowing for an expression Akaike's information criterion directly in terms of deviance:

$$AIC = 2k + D$$

For models with sufficiently similar test suite results, the model with the smallest AIC is always chosen. Continual re-evaluation of optimal models across keywords ensures that models stay abreast of market changes.

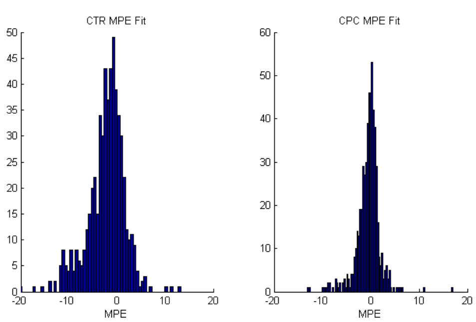

Figure 4: Fit Mean Percent Error

5.1 Fit and Prediction

This section shows the fit and prediction results for CTR and CPC in terms of mean percent error (MPE) in Figure 4 and Figure 5. The mean and standard deviation of fit and prediction results for CTR and CPC are shown in Table 2 and Table 3. The results reflect the modeling of a group of high volume keywords. Model 2 (M2) was picked for modeling both the CTR and CPC in these cases as it is the intersection of a minimization of the absolute value of both the mean and standard deviation of the residuals and also the AIC.

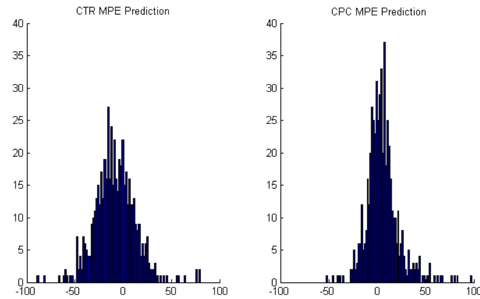

Figure 5: Prediction Mean Percent Error

5.2 Discussion

Leveraging hierarchical structure in keyword response data reduces variance in SEM metric modeling. Variance reduction translates into improvement in profit at stochastic optimum. Production system rules-based bidding and periodic re-evaluation of models all play a part in a successful SEM bid. Validation and measuring the response of new models in the real world remains difficult (requiring holdout test data) because the testing of bids can only occur in a live system as most search engines use proprietary ranking algorithms.

Many open questions still exist such as what other predictors are effective? How can one incorporate competitive effects? Is endogeneity introduced by search performance ranking? Also, including demographics, psychographics and personalization make data sparse, but volatility is also reduced.

REFERENCES

Akaike, H. (1974), "A New Look at The Statistical Model Identification," *IEEE Transactions on Automatic Control 19 (6)*: 716723.

McCullagh, P. and Nelder, J.A. (1989), "Generalized Linear Models," Chapman and Hall, London.

Nelder, J. A. and Wedderburn, R. W. M (1972), "Generalized Linear Models," *Journal of the Royal Statistical Society*.

Search Marketing Benchmark Guide 2007, Marketing Sherpa, Warren, Rhode Island.