Position Effects in Web Search Click Behavior

Charles Mathis
Google, Inc.
1600 Ampitheater Parkway
Mountain View, CA 94043
cmathis@google.com

ABSTRACT

When a search engine presents a set of document links, the rates and characteristics of clicks on each link may be influenced by display effects. Users may be more predisposed to click on some positions than others, whether due to attention and reading order, their interpretation of the position as additional relevance information, or other reasons. We quantify this effect, and fit models for the influence of these position effects on clickthrough. We also fit empirical prior distributions to clickthrough rates, and use posterior estimation as a smoother of empirical averages.

Categories and Subject Descriptors

G.3 [**Mathematics of Computing**]: Probability and Statistics—*correlation and regression analysis, distribution functions*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*relevance feedback*

General Terms

Measurement, relevance feedback, statistics

Keywords

Measurement, relevance feedback, statistics

1. INTRODUCTION

When a user issues a query to a search engine, they are typically given a list of document summaries, which usually consist of a title linking to the page and a sample of text from the page.

Users' probability of clicking on particular document links is potentinally affected not only by the utility of the result as percieved solely as a function of its title, URL, and text summary, but also by the result summary's position in the list of results. Users may never read many of the summaries, and they may be predisposed to use the search engine's ordering as an indicator of relevance.

We use aggregated, anonymized data on historical searches to formulate generative and predictive models of click probabilities. In section 2, we find that the data is consistent with a beta-binomial model, where the true clickthrough probabilities of the population of results at a given position have a beta distribution. Given a value of this prior, and a certain number of impressions (views) of a result, the number of clicks is binomial distributed. In section 3, we formulate a joint distribution of clickthrough probabilities for a given result across multiple positions, which is consistent with the beta marginal distributions. In section 4, we find that this distribution accurately describes clickthrough data, and yields a predictive model for clickthrough at one position given the empirical clickthrough rate at another. This algebraically simple model is as accurate as more complex and arbitrarily flexible regression techniques, such as cubic splines and kernel-smoothed moving averages.

These effects have also been studied for the commercial Yahoo! search engine [2, 3, 1] and a Cornell University academic search engine [5, 6], using logs data to evaluate search engines and their results, and controlling for the position effect by randomization or by inclusion as a model feature. Another tack is the use of eye-tracking studies [4] to measure the distribution of attention across positions and click probabilities conditional on attention.

WWW2009, April 20-24, 2009, Madrid, Spain.
.

2. SMOOTHING CLICKTHROUGH RATES

In this section we study the estimation of clickthrough rates at a single display position. This will be a component of the cross-position estimation of later sections, and will motivate a joint distribution for the position effect model.

The empirical clickthrough rate of a result, the number of clicks (ignoring duplicate clicks on the same result for the same query) divided by the number of impressions or views of the result at that position, is not necessarily the best estimator of the true clickthrough probabality of the result. If a query is observed once, each result will have empirical clickthrough of 0 or 1, which is unrealistic as a long-term rate. We wish to better estimate clickthrough probability of a result at a given position before using it for other purposes, such as projecting the result's clickthrough probability at another position.

Our knowledge about clickthrough rates across all results and queries informs estimates for individual results. An empirical clickthrough rate's distance above or below average is likely part sampling noise and part true signal. We can improve our estimate of this true clickthrough probability by regressing to the mean of the population for that position.

Knowing a prior distribution of clickthrough probabilities allows derivation and use of estimates based on the posterior distribution given the observed clicks and impressions.

The distribution of the number of clicks given the prior clickthrough probability is taken to be binomial, as we only count one query and one click per positon per user, and we consider the users independent (despite the possibility of one user issuing the same query from multiple computers or mobile devices, for example).

We desire to fit a prior from data rather than choosing one arbitrarily. A practical difficulty with estimating such a prior is that true clickthrouh probabilities are never observed. The distribution of empirical clickthrough rates has different properties, such as point masses and a larger variance. If clickthrough had a uniform prior and we observed only a single impression per result, the variance of clickthrough rates would be that of a fair coin, 1/4, whereas the variance of the uniform prior is 1/12.

The uniform is a special case of the beta distribution, which is the conjugate prior to the binomial distribution and fits clickthrough data with surprising accuracy. If the prior of a unit interval valued random variable p is beta distributed with parameters (α, β), and we then observe a random variable X which is binomial distributed with parameters (n, p), the posterior distribution of p given X will be beta $(\alpha + X, \beta + n)$. The posterior mean is then

$$\frac{X+\alpha}{n+\alpha+\beta} = \frac{X}{n}\frac{n}{n+\alpha+\beta} + \frac{\alpha}{\alpha+\beta}\frac{\alpha+\beta}{n+\alpha+\beta}$$

which can be interpreted as a linear shrinkage of the empirical mean $\frac{X}{n}$ partway to the population mean $\frac{\alpha}{\alpha+\beta}$, with less shrinkage when the sample size n is larger and the empirical mean more accurate.

We choose the beta prior parameters (α, β) via the method of moments, matching the mean and variance of the model to the empirical distribution. Rather than equating the empirical mean and variance of clickthrough rates to the beta directly, we equate them to the beta-binomial distribution of clickthrough rates. Otherwise the variance will be overestimated, as in the uniform example above.

We use the method of moments rather than maximum likelihood because it is computationally simpler, less outlier sensitive, and unbiased (the model and empirical means match, leading to accurate predictions of population aggregate clickthrough).

Denote the clickthrough rates as $Y_i = \frac{X_i}{n_i}$ for $i = 1$ to N. Each X_i is distributed binomial (n_i, p_i) where the number of impressions n_i is observed and the unobserved p_i are each drawn independently from the same beta (α, β) prior.

Recall that a beta distribution has support on the unit interval $(0, 1)$, with density function $p(x|\alpha, \beta) = \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}$, where B is the beta integral function $B(\alpha, \beta) = \int_0^1 p^{\alpha-1}(1-p)^{\beta-1}dp$.

The mean of the clickthrough rates and underlying click probabilites are the same:

$$EY_i = \int E[Y_i|p]dP(p) = \int p \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}dp$$
$$= B(\alpha+1, \beta)/B(\alpha, \beta) = \frac{\alpha}{\alpha+\beta}$$

The second moment is

$$EY_i^2 = \int E(Y^2|p) \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}dp$$
$$= \int (p^2 + \frac{1}{n_i}p(1-p)) \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}dp$$
$$= (1 - \frac{1}{n_i})Ep^2 + \frac{1}{n_i}Ep$$
$$= (1 - \frac{1}{n_i})\frac{(\alpha+1)}{(\alpha+\beta+1)}\frac{\alpha}{\alpha+\beta} + \frac{1}{n_i}\frac{\alpha}{\alpha+\beta}$$

Abbreviate the empirical averages

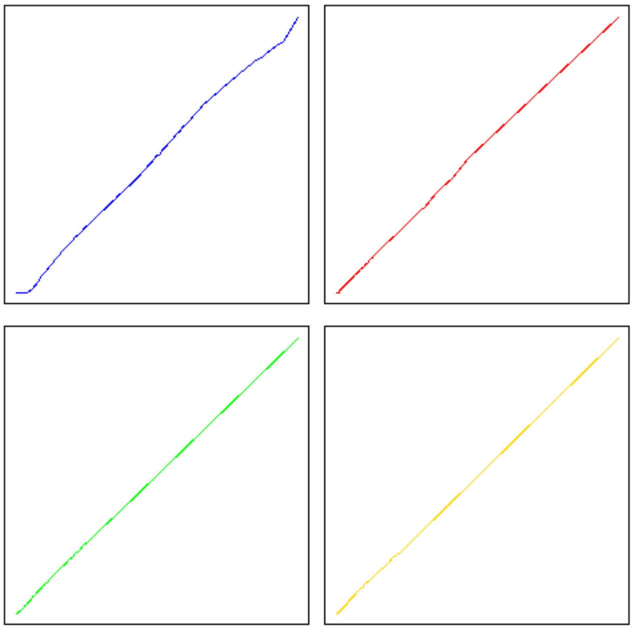

Figure 1: Quantile-quantile plots for first four positions, beta-binomial model of clickthrough vs empirical

$$\mu = \frac{1}{N}\sum_{i=1}^N Y_i \qquad \nu = \frac{1}{N}\sum_{i=1}^N Y_i^2 \qquad \zeta = \frac{1}{N}\sum_{i=1}^N \frac{1}{n_i}$$

We equate the observed moments to those of the model:

$$\mu = \frac{\alpha}{\alpha+\beta} \qquad \nu = \zeta\mu + (1-\zeta)\frac{(\alpha+1)}{(\alpha+\beta+1)}\mu$$

The method of moments estimates then arise from solving these equations for α and β:

$$\alpha = \mu(\frac{\mu(1-\mu)(1-\zeta)}{\nu-\zeta\mu-(1-\zeta)\mu^2} - 1) \qquad \beta = (1-\mu)(\frac{\mu(1-\mu)(1-\zeta)}{\nu-\zeta\mu-(1-\zeta)\mu^2} - 1)$$

Figure 1 shows four quantile-quantile plots, one for each of the first four result positions. The X axes are the quantiles of the beta-binomial distributions fit with the method of moments, and the Y axes are quantiles of the empirical clickthrough rates from a sample of 10,000 results at each position. The parameters were fit from separate training sets of 10,000 results per position.

A quantile-quantile plot will be a straight line through points $(0, 0)$ and $(1, 1)$ iff the two distributions are the same up to scalings and shifts (which in this case are fixed, as both distributions range from 0 to 1). Each of these curves has a correlation greater than .999. Even if the data arose from the model distribution exactly, the lines will not be perfectly straight, due to sampling variance. The quantile-quantile correlations between bootstrap resamplings of the clickthrough data had lower correlations for almost half the resamplings, indicating that the error in the beta-binomial is within the sampling variation of our dataset.

We can evaluate the usefulness of the posterior estimate using these priors by using it to predict clickthrough rates on test data from training data for the same results, and comparing the error to that using empirical fractions as predictions. We made a random split of the aforementioned test sets, in this case by splitting the individual impressions of each result. For example, a result with 3 impressions might have 1 impression in the training set and 2 impresions in the test set.

We computed empirical and posterior clickthrough rates for each result from the training half, and measured the L1 and L2 error between these and the empirical clickthrough rate of the same result on the test half. We averaged these errors across results in two ways, uniformly with respect to each unique result, and weighed with respect to the number of impressions. The latter represents the expected error on a result drawn uniformly from web search traffic.

The posterior mean estimator had 20% to 34% lower error than the empirical mean estimator, ranging across the possible combinations of position, error metric, and weighting.

3. A POSITION EFFECT MODEL

When transforming interval-valued clickthrough rates to real-line valued variables via link functions such as logit and probit, we do not observe elliptic-shaped data (as we would with multivariate Gaussian data on this scale, for example). The distribution is assymetric and tends to stay in the lower triangle where clickthrough at lower display positions is lower than at higher display positions.

We consider a generative model for clickthrough across multiple positions, finding a joint distribution whose marginals are consistent with the beta distributions found for individual positions. This leads to a choice of an algebraically simple prediction curve which performs as well as much higher-dimensional parametric or nonparametric curves, such as splines and kernel-smoothed moving averages, and outperforms regression on the ordinary, logarithmic, or logistic scales.

The Dirichlet distribution is a well known joint distribution with beta marginals. When a set of independent gamma random variables are divided by their sum, the resulting marginals are beta distributed, and always sum to 1.

The correlation between clickthroughs for a result across multiple positions is likely positive, and need not satisfy any sum constraint exactly. Nevertheless, we can use the same observation that a beta distribution arises from a ratio of gammas.

Let $X_0, X_1, X_2, Y_0, Y_1, Y_2$ be gamma distributioned random variables with the same scale parameter, say 1 without loss of generality, and shape parameters $a_0, a_1, a_2, b_0, b_1.b_2$. Recall that a gamma distribution with scale parameter θ and shape parameter k has density $x^{k-1}e^{-x/\theta}\theta^{-k}/\Gamma(k)$, supported on the positive real line, where $\Gamma(k) = \int_0^\infty t^{k-1}e^{-t}dt$.

The sum of gammas with the same shape parameter are gamma, and the ratio of a gamma to its sum with another independent gamma is beta. So

$$B_i = \frac{X_0 + X_i}{X_0 + X_i + Y_0 + Y_i}$$

are beta distributed, for $i = 1, 2$. In general B_1 and B_2 are nonnegatively correlated, and X_0, Y_0 reflect their dependence.

Conditional on B_1, the shared random variables X_0, Y_0 are each distributed as the product of a beta and a gamma random variable: $X_0 = B_1U_1Z_1$ and $Y_0 = (1 - B_1)V_1Z_1$, where U_1 is beta (a_0, a_1), V_1 is beta (b_0, b_1), and Z_1 is gamma $(a_0 + a_1 + b_0 + b_1)$, all independent. In fact Z_1 is the denominator $X_0 + X_1 + Y_0 + Y_1$ of B_1, which is independent of B_1. These conditional distributions arise from this independence: given B_1, $X_0 + X_1$ is distributed as B_1Z_1, and then the ratio $U_1 = \frac{X_0}{X_0+X_1}$ is beta (a_0, a_1) independently. The distribution for Y_0 follows from analogous considerations of $1 - B_1$.

Meanwhile, X_2 and Y_2 are independent of B_1. These conditional distributions allow an estimate of B_2 given B_1 which has the parameteric form of a first-order rational function

$$f(B_1) = \frac{c_1B_1 + c_2}{d_1B_1 + d2}$$

where

$$c_1 = (a_0 + a_1 + b_0 + b_1)\frac{a_0}{a_0+a_1}$$
$$c_2 = a_2$$
$$d_1 = (a_0 + a_1 + b_0 + b_1)(\frac{a_0}{a_0+a_1} - \frac{b_0}{b_0+b_1})$$
$$d_2 = a_2 + b_2 + (a_0 + a_1 + b_0 + b_1)\frac{b_0}{b_0+b_1}$$

4. EMPIRICAL FITTING

Figure 2 is a quantile-quantile plot of the model (X axis) vs the empirical (Y axis) distributions of the ratio of a result's clickthrough at second position to its clickthrough at first position. The model is the joint-beta distribution from section 3, which we have already seen has consistent marginals with the data. We find, as in this ratio example, that the joint distribution is consistent from the data as well.

Fitting the marginal distributions as in section 2 provides four parameter constraints, $\alpha_i = a_0 + a_i, \beta_i = b_0 + b_i$, for $i = 1, 2$, leaving two free parameters. Extending the method of moments to equate the empirical and model covariance provides one additional constraint.

Also, positions are ordered in nature. Except for edge effects at the bottom of a page, lower positions lead to lower clickthrough. We can constrain the joint distribution thusly.

In the first-order rational function predictor, the condition that a higher position yield higher clickthrough can be formulated as $f(B_1) \leq B_1$ if $f(B_1)$ represents clickthrough at a lower position, or $f(B_1) \geq B_1$ if a higher position. The former case implies $0 = f(0) = \frac{c_2}{d_2}$, so $c_2 = a_2 = 0$ and $a_0 = \alpha_2$. The latter case similarly implies $b_2 = 0$.

Figure 3 shows a scatterplot of posterior clickthrough rates for results observed at both the first (X axis) and second (Y axis) positions. As expected, almost all the points lie in the lower triangle, and the amount of mass in the upper triangle is consistent with the binomial sampling variation. Using posterior clickthrough estimates, as derived in section 2, leads to better prediction than using empirical click to impression ratios.

The straight lines in figure 3 represent linear fits of clickthrough at the second position y given the clickthrough at the first position x, one fit to minimize expected squared error, and the other fit to minimize expected squared error on the logarithmic scale. Note that such linear models $y = \frac{c_1}{d_1}x$ are special cases of the above first-order rational function model, with $c_2 = d_2 = 0$.

A logistic shift model $\frac{y}{1-y} = c\frac{x}{1-x}$ is also a special case of the first-order rational function, and maps both 0 to 0 and 1 to 1. Logarithmic or logistic models which scale as well as shift, i.e. $l(y) = s \cdot l(x) + r$ where $l(x) = \log(x)$ or $l(x) = \log \frac{x}{1-x}$, will not obey the rules $f(x) <= x$ or $f(x) >= x$ unless the slope is 1.

In figure 3, the green, monotone curved line represents the fit of first-order rational function. The wavy blue line is a kernel estimate, or weighted moving average.

f tracks the kernel estimate closely in the left half of the graph, where the data is denser. In the right half, the moving average oscillates, indicating overfitting. A kernel which adjusts its bandwidth based on the data density, or uniformly weights the k nearest neighbors, can be tuned so as not to oscillate as much. Given enough data, the kernel regression, or a cubic spline with an increasing number of knots, will converge to the true conditional mean of the response given the predictor, and can thus be thought of as a noisy ground truth.

When using a training and test data split, the first-order rational model has slightly lower absolute or squared error than kernel estimates or cubic splines with many more parameters. We interpret this to mean that the model is very predictively accurate given its simplicity, and by construction is consistent with the generative distributions of clickthrough rates.

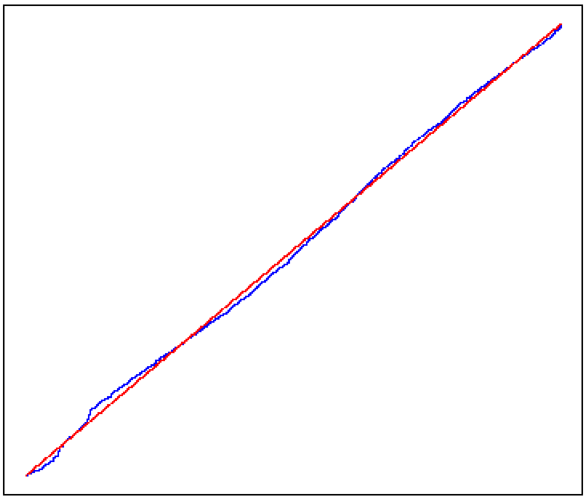

Figure 2: Quantile-quantile plot of clickthrough ratio between second and first positions, model vs empirical distributions

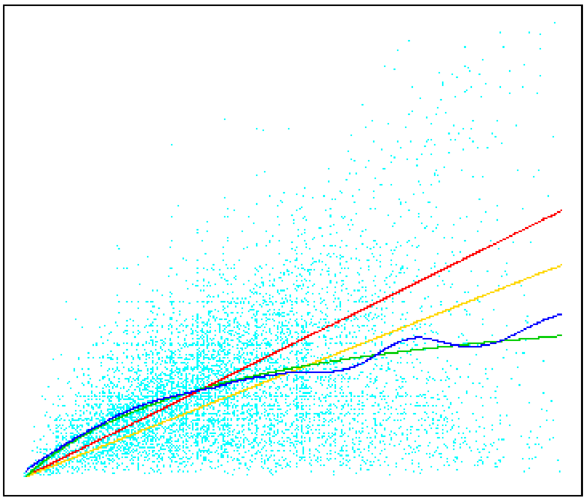

Figure 3: Clickthrough at first vs second positions

5. DISCUSSION

We have seen that the distributional properties of clickthrough rates are amenable to classical parametric models, which motivate prediction rules as accurate as other higher-parameter or nonparametric models.

Even with the correct distributional knowledge, prediction of clickthrough for a given result and position has a certain natural variance. The dispersion of points in figure 3 is greater than that from the binomial variation alone.

A general clickthrough prediction model can use other features about the query, the result, and the other displayed results (such as their clickthrough) to reduce the remaining variation. These features can be thought of as explanatory variables for the variation in the beta prior of clickthrough.

6. REFERENCES

[1] G. Dupret, V. Murdock, and B. Piwowarski. Web search engine evaluation using clickthrough data and a user model. In *Proceedings of WWW, workshop on query log analysis*, 2007.

[2] G. Dupret, B. Piwowarski, C. Hurtado, and M. Mendoza. A statistical model of query log generation. In *Proceedings of SPIRE*, 2006.

[3] G. Dupret, B. Piwowarski, and V. Murdock. A bayesian model of query log generation. In *Proceedings of SIGIR*, 2007.

[4] L. Granka, T. Joachims, and G. Gray. Eye-tracking analysis of user behavior in www search. In *Proceedings of SIGIR*, 2004.

[5] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of SIGIR*, 2005.

[6] F. Radlinkski and T. Joachims. Minimally invasive randomization for collecting unbiased preferences from clickthrough logs. In *Proceedings of AAAI*, 2006.