

网络广告关键词选择优化

王 非

WANG Fei

广东外语外贸大学 信息科学技术学院, 广州 510006

School of Informatics, Guangdong University of Foreign Studies, Guangzhou 510006, China

E-mail: gdufs.wangfei@gmail.com

WANG Fei. On keyword selecting for web advertising. Computer Engineering and Applications, 2009, 45(7): 227-229.

Abstract: A key task for a search engine advertising system is to determine what advertisements should be displayed, and in what order, for each query received. This is a gaming process between advertisers and search engine: the engine needs to accurately estimate the click-through rate of advertisements, improves its revenue and satisfaction of advertisers and users. Advertisers need to guess engine ranking model, bid advertising slots, adjust strategies according to the advertising effects and perform other search engine optimizing actions. This paper guesses and conducts a common CPC mode search engine advertisements ranking method, and applies this method on guilding advertisers' keyword selecting strategy.

Key words: web advertising; search engine optimizing; pay for searching

摘 要: 搜索引擎广告系统的核心任务是针对用户提交的每个查询, 决定以何种顺序显示哪些广告。这是广告商和搜索引擎之间的一个博弈过程: 搜索引擎需要准确地估算广告的点击率, 提高广告收益、广告商的满意度及用户的满意度; 广告商则需要猜测搜索引擎的排序模型、竞价广告栏并依据广告效果调整广告策略以及执行其他的搜索引擎优化动作。探讨了推测 CPC 模式下搜索引擎广告排序模型的一般方法, 并据此指导广告商的关键词选择策略。

关键词: 网络广告; 搜索引擎优化; 付费搜索

DOI: 10.3778/j.issn.1002-8331.2009.07.069 **文章编号:** 1002-8331(2009)07-0227-03 **文献标识码:** A **中图分类号:** TP375

1 前言

广告收入是绝大多数现代搜索引擎的主要资金来源^[1]。网络广告有多种形式, 主要有:

(1) CPC (Cost-Per-Click), 广告商按照广告链接被点击次数向搜索引擎支付一定费用, 该方式中是目前最常用的方式, Google、Yahoo 等都采用 CPC 方式。

(2) CPI (Cost-Per-Impression), 广告商按照广告链接展示时间的长短向搜索引擎付费。

(3) CPA (Cost-Per-Action), 只有当用户采取了某种广告商期望的动作, 如购买商品, 广告商才向搜索引擎支付费用。

搜索引擎广告系统的核心任务是针对提交的每个查询, 决定以何种顺序显示哪些广告。通常, 广告商会指定什么情况下其广告应该出现(如, 当查询中包含某些特定用语时), 因此很多情况下搜索引擎只需要排序广告链接即可。研究^[2]表明, 用户点击广告链接的概率随广告显示位置的不同变化非常大, 如图 1 所示, 随着显示位置的下降, 其点击概率直线下降。因此, 将收益最高的广告放在第一位对搜索引擎是最有利的。鉴于此点, 准确地估计广告的点击率, 优化广告链接排序对搜索引擎的收益至关重要。比如, 某广告点击一次付费 1 元, 那么被点击 100 次才会带来 100 元收益, 假如搜索引擎过高的估计了该广

告的点击率, 该广告的排序会不合理的提升, 挤占广告位, 降低高收益广告链接的流量, 直接影响搜索引擎的收益, 低估广告点击率同样如此。

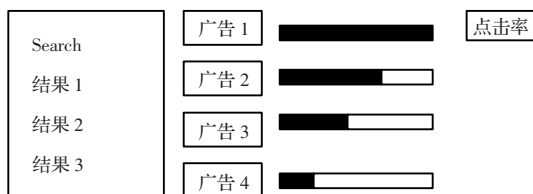


图 1 广告点击率变化示意图

随着更多的广告商进入网络广告市场, 同时现存广告商不断加大广告力度, 搜索引擎的优化问题日渐凸显: 广告商需要猜测搜索引擎的排序模型, 结合广告效果动态调整其广告策略(如变化广告描述用语等), 以保证自己的广告投入物有所值; 另一方面, 搜索引擎则需要不断的针对广告商的调整重新估计点击率, 在稳定广告商满意度及引擎用户满意度的同时, 保证自己收益的最大化。本文将探讨推测 CPC 模式下搜索引擎广告链接排序的一般方法, 并据此指导广告商的网络广告关键词选择策略。

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.70473066)。

作者简介: 王非(1976-), 男, 博士, 讲师, 主要研究领域为信息检索、电子商务、知识管理。

收稿日期: 2008-01-21 **修回日期:** 2008-03-28

2 搜索引擎预期收益估算

通常,搜索引擎接受的每个查询可用的广告数量会远远超过可用广告栏的数目。但绝大多数用户只会浏览查询结果的第一页^[3],而搜索引擎每页的广告栏数目一般只有3~8个,因此为了优化广告质量(以用户点击率作标准),最大化广告收益,搜索引擎基本上都依据预期收益 $E(X, Y)$ 排序广告:

$$E(X, Y) = CPC(X, Y) * C(X, Y) \quad (1)$$

公式(1)表示对搜索关键词 X , 搜索引擎在广告栏位置 Y 上的预期广告收益。其中: $CPC(X, Y)$ 代表广告链接被点击一次, 标中广告栏位置 Y 的厂商需要向搜索引擎支付多少费用; $CTR(X, Y)$ 代表在广告栏位置 Y 上广告链接的预期被点击次数。如果能具体化公式(1), 就能推测对特定关键词 X , 搜索引擎排序广告位的算法, 进而构建搜索引擎收益模型, 指导厂商的关键词选择策略。

2.1 CPC 估算

网络广告位置同访问流量增加之间的正比关系使得广告商竞相追逐更好的广告位置, 但广告链结会使搜索结果片面化, 出于维持和提高用户满意度的考虑, 搜索引擎通常会严格控制分配广告栏。目前, 网络广告栏的分配机制主要有四种:

(1) v 模式, 即 Vickrey auction 模式, 也称 second price auction 模式或第二价格竞拍机制。简单讲, 该模式就是出价最高者实际以第二名的出价取得标的, 其主要代表为被 yahoo 收购之前的 overture, 在 overture 的实现中, 广告商通过一个竞拍代理竞拍, 该代理会自动在最接近竞争者报价基础上加价 1 分, 以这种方式模拟第二价格竞拍机制。

(2) r 模式。该模式强调广告链接的相关性, 关注搜索引擎用户的满意度, 相关性高的广告取得较好的广告栏位置。

(3) PPR (Posted Price Ranking) 模式, 搜索引擎直接给出广告位的报价, 无需竞买, 先来先得, 有点类似于报纸等传统媒体销售广告位的方法。信息门户及类似站点销售广告位置常用这种方式,

(4) v & r 模式, 结合 v 模式和 r 模式, 其主要代表是 Google。Google 将广告栏位置的点击历史同 v 模式相结合设计广告栏分配机制, 兼顾了广告栏收益和用户满意度。

考虑检索词 X 的广告栏分配: 假设有 s 个公司希望竞标 X 的广告, 搜索引擎提供了 k 个广告位。

令 v_j 代表第 j 个广告商的 CPC 报价, r_j 代表其广告的真实相关度(未经搜索引擎提升排名), 有 $0 < v_j, 0 < r_j < 1$ 。令 $A: I \rightarrow J$ 为搜索引擎中将广告位 i 分配给广告商 j 的排序函数。

v 模式中, 搜索引擎选择出价最高的 k 个广告商并顺序排序他们的广告, 故函数 $A(i+1) = [j | v_j = \max(v_m), m \in [1, s]]$ 且 $m \neq A(t), t \in [0, i], A(0) = 0$, 而广告位 i 的实际价格为投拍广告位 $i+1$ 的广告商的 CPC 报价, 即 $CPC(X, i) = v_{A(i+1)}$ 。

r 模式中, 搜索引擎选择出价最高的 k 个广告商并按照其广告的预期点击率排序广告, 故函数 $A(i+1) = [j | r_j = \max(r_m), m \in [1, s]]$ 且 $m \neq A(i), A(0) = 0$, 而广告位 i 的实际价格为出价 $k+1$ 高的广告商的 CPC 报价, 即 $CPC(X, i) = v_{A(k+1)}$ 。

PPR 模式中, 搜索引擎依时间段为每个广告位设定一个底价, 竞标过程中, 从出价最高的 k 个广告商中选则所有出价超过底价的广告商。

v & r 模式中, 搜索引擎选择出价最高的 k 个广告商并顺序

排序他们的广告, 定价方式是标准第二价格竞拍机制的变种。如果 $v_{A(i)} > v_{A(i+1)}$, 即赢得广告位 i 的厂商出价高于赢得广告位 $i+1$ 的厂商, 则 $CPC(X, i) = v_{A(i)}$; 否则:

$$CPC(X, i) = v_{A(z)}, \{z = \min(t) | v_{A(t)} * r_{A(t)} \geq v_{A(i+1)} * r_{A(i+1)}\}$$

2.2 C 估算

为了最大化收益及广告商满意度, 搜索引擎的点击付费系统必须能预测用户对广告链接可能的动作, 最大化其点击的概率。

研究^[2]表明, 即使具有相同关键词的广告位, 其点击率的变化仍然非常大, 甚至可以相差十倍以上。这说明除广告本身相关度外, 检索情境中有某些因素在影响用户行为。

(1) 广告的点击率依赖于广告本身的相关性及广告栏的位置。一般来讲, 用户倾向于点击排序位置高的广告栏。文献^[3]曾提出一个关注衰减指数模型, Overture 2003 年的数据分析也证明了该模型, 即广告栏位置的被关注程度呈指数衰减。借鉴此模型思路, 假设广告栏 i 处广告的无干扰基本点击率可用公式 $r_{A(i)} / \theta^{i-1}$ 计算 (θ 为关注衰减常数)。

(2) 通常情况下, 搜索引擎用户会区别对待搜索结果中的正常链接和广告链接, 以参数 b ($b \geq 1, b$ 越大表示用户越看重正常结果) 表示用户的这种态度。显然, 这种偏好会影响广告点击率。

(3) 广告栏中其它广告的相关度水平无疑会影响用户对广告的兴趣, 进而影响点击率。为了简化分析, 用 k 个广告栏的平均真实相关度水平 ($\sum_i r_{A(i)} / k$) 表示这种影响。较低的平均相关度水平肯定会导致用户对广告栏的忽视行为, 而较高的平均相关度水平则会增加用户点击广告的可能性。该因素还会加强/削弱用户对正常链接/广告链接的偏好。

结合上述因素, 使用公式(2)猜测搜索引擎的点击率估算方法:

$$CTR(X, Y) = \frac{1}{1 + e^{-\lambda}} \quad (2)$$

$$\lambda = (w_1 * (r_{A(Y)} / \theta^{Y-1}) + w_2 * (\sum_i r_{A(i)} / k))^b$$

其中, w_1 和 w_2 分别为搜索引擎对因素 1 和因素 3 的加权, θ 为关注衰减常数, 这里取 $\theta = 1.428$, 该常数已在多个实测数据集上得到验证^[4]。那么就有:

$$C(X, Y) = \int_0^1 CTR(X, Y) dr_{A(Y)} \quad (3)$$

2.3 收益模型

考虑采用 v & r 广告栏分配模式(其它分配模式的收益模型可类似得出)的搜索引擎 E , 其 CPC 函数可定义为:

$$CPC(X, Y) = \begin{cases} \text{若 } v_{A(Y)} > v_{A(Y+1)}, Y \in [1, k]; \text{ 则为 } v_{A(Y+1)} \\ \text{若 } v_{A(Y)} \leq v_{A(Y+1)}, Y \in [1, k]; \text{ 则为 } v_{A(Z)} \end{cases}$$

其中, $A(i+1) = [j | v_j = \max(v_m), m \in [1, s]]$ 且 $m \neq A(t), t \in [0, i], A(0) = 0$, $z = \{\min(t) | v_{A(t)} * r_{A(t)} \geq v_{A(i+1)} * r_{A(i+1)}, t \in [1, k]\}$ 。

E 对关键词 X 的总体收益模型猜测如下:

$$E(X) = \sum_{Y=1}^k CPC(X, Y) * C(X, Y) = \sum_{Y=1}^k CPC(X, Y) * \int_0^1 CTR(X, Y) dr_{A(Y)} \quad (4)$$

3 关键词选择优化

一般来讲,厂商 P 的竞价原则是:在既定经费范围 F 内,最大化预期点击量 $C(Y)$ 。

考虑 v&r 模式下(其它分配模式的竞价优化模型可类似得出), S 个厂商竞价搜索引擎 E 提供的关键词 T 的 k 个广告位。假设厂商 P 标中编号 H 的广告栏,那么其预期广告支出为:

$$F_E(T,H)=CPC(T,H)*C(T,H)=$$

$$CPC(T,H)*\int_0^1 CTR(T,H)dr_{A(H)}=$$

$$CPC(T,H)*\int_0^1 \frac{1}{1+e^{-(w_1*(r_{A(H)}^{\theta^{H-1}})+w_2*(\sum_{i=1}^k r_{A(i)/k})^{\theta})}} dr_{A(H)}=$$

$$\begin{cases} v_{A(H+1)} * \int_0^1 \frac{1}{1+e^{-(w_1*(r_{A(H)}^{\theta^{H-1}})+w_2*(\sum_{i=1}^k r_{A(i)/k})^{\theta})}} dr_{A(H)}, \text{若 } v_{A(H)} > v_{A(H+1)} \\ v_{A(z)} * \int_0^1 \frac{1}{1+e^{-(w_1*(r_{A(H)}^{\theta^{H-1}})+w_2*(\sum_{i=1}^k r_{A(i)/k})^{\theta})}} dr_{A(H)}, \text{若 } v_{A(H)} \leq v_{A(H+1)} \end{cases} \quad (5)$$

其中: $A(i+1)=j|v_j=\max(v_m), m \in [1, s]$ 且 $m \neq A(i), i \in [0, i], A(0)=0, z=\{\min(t)|v_{A(t)}*r_{A(t)} \geq v_{A(H+1)}*r_{A(H+1)}, t \in [1, k]\}$ 。

搜索引擎返回结果的排序是 r 值的体现,故可借此推测公式(5)中的 $r_{A(H)}$ 值:搜索关键词 T ,取其前 100 个返回结果,按照排序位置 i 赋予相关值 $o_i=101-i$,那么 $r_{A(H)}$ 可以估算如下:

$$r_{A(H)} = \sum_{i=1}^{100} r_{A(H)}(i)/5050 \quad (6)$$

$$r_{A(H)}(i) = \begin{cases} \text{如果厂商 } A(H) \text{ 出现在结果位置 } i \text{ 上,为 } o_i \\ \text{否则,为 } 0 \end{cases}$$

因此,对特定关键词 T ,一旦确定目前 CPC 出价最高的 k 个厂商,就可以依据公式(5)估算他们的广告预期点击量以及广告预期花费。

目前,采用 v&r 模式的搜索引擎主要是 Google, Overture 在被 Yahoo 收购后也考虑转向 v&r 模式。Google 的 adwords 系统并不提供竞价过程的详细信息,厂商无法得知自己的实际 CPC 值也无法得知竞争对手的报价。不过,仍然可以估算某些情况下的 CPC 值。

以 Google 为例,如果关键词 T 不是全新的广告词,即 T 已经在 adwords 系统中使用过一段时间,那么在执行搜索时,结果页面的右侧应该会出现一些广告。前文中,在估算 $r_A(H)$ 时取了 100 个结果,Google 中,这些结果通常分 10 页显示,可以认为这 10 页结果中出现的所有广告厂商就是 adwords 中对关键词 T 出价最高的 k 个厂商,这也是 A 函数取值的体现。一旦确定了这 k 个厂商以及他们同广告位之间的对应关系 $A(H)$,就可以依据公式(6)推测全部 $r_{A(H)}$,依据公式(3)推测所有 k 个广告位的点击量 $C(T,H)$ 。

再者,adwords 系统采用 v&r 模式,因此有:

约束 1 $v_{A(1)}*r_{A(1)} \geq v_{A(2)}*r_{A(2)} \geq \dots \geq v_{A(k)}*r_{A(k)}$

约束 2 $v_{A(1)}*C(T,1) \geq v_{A(2)}*C(T,2) \geq \dots \geq v_{A(k)}*C(T,k)$

同时,adwords 系统提供了一个点击量估算工具,估算结果

平均每次点击费用: **C4.37** 最高每次点击费用为 **¥14.22**
估算的每日点击次数: 652 · 819 (每日预算为 ¥4,100.00)

最高每次点击费用:	每日预算	获得新的估算值			
关键字	搜索量	估算平均每次点击费用	估算的广告排名	估算的每天点击量	估算的每天费用
旅游		¥3.50 - ¥4.96	1 - 3	652 · 819	¥2,300 - ¥4,100
搜索网络总计		¥3.50 - ¥4.96	1 - 3	652 · 819	¥2,300 - ¥4,100

中,平均 CPC 值和最大 CPC 值可以作为约束条件 3 和约束条件 4。图 2 为关键词“旅游”的估算结果。

约束 3 $v_{A(1)}+v_{A(2)}+\dots+v_{A(k)}=\text{平均点击费用} * k$

约束 4 $v_{A(i)} \leq \text{最高点击费用}$

在 adwords 系统中, k 一般为 3,由于 $r_{A(H)}, C(T,H)$ 可估算得出,平均点击费用、最高点击费用可查询得知,那么这四个约束必定可解出 $v_{A(H)}$,从而得出 CPC 值。

至此,公式(5)中的全部变量都估算得出。于是,可以借助公式(5)优化厂商 P 的关键词选择策略,体现在两个方面:

(1)单关键词选择优化。如果厂商 P 只准备采用单个广告关键词,那么可以:

①在竞价过程中,跟踪最高的 k 个厂商报价,随时引用公式(5)计算自己所竞标广告位的预期广告费用 $F_E(T,H)$,考察 $F_E(T,H)$ 是否会超出既定经费额度 F ,依此决定是否继续竞标广告词 T 或是否调低预竞标广告位。

②同时竞标多个关键词,比较在预期广告费用近似的情况下,各个关键词的 C 值,选择其中最高者确认竞标。

(2)多关键词选择优化。通常情况下,厂商会选择一组广告关键词^[5],那么相应的优化策略涉及两个方面:

①如何确定这一组关键词?

②广告费用的分摊。CPC 和 C 是同步增长的,因此确定每个关键词的 CPC 上限很重要,广告费用的分摊过程实质上就是厂商所有广告关键词总体 C 值最大化的求解过程。

可以就当前厂商经营领域最常用的一批 n 个关键词生成两个矩阵,即点击量矩阵 M_C 和点击费用矩阵 M_{CPC} ,如表 1 和表 2。

表 1 点击量矩阵 M_C

关键词	广告位			
	1	2	...	k
T_1	$C(T_1,1)$	$C(T_1,2)$...	$C(T_1,k)$
T_2	$C(T_2,1)$	$C(T_2,2)$...	$C(T_2,k)$
\vdots	\vdots	\vdots		\vdots
T_n	$C(T_n,1)$	$C(T_n,2)$...	$C(T_n,k)$

表 2 点击费用矩阵 M_{CPC}

关键词	广告位			
	1	2	...	k
T_1	$CPC(T_1,1)$	$CPC(T_1,2)$...	$CPC(T_1,k)$
T_2	$CPC(T_2,1)$	$CPC(T_2,2)$...	$CPC(T_2,k)$
\vdots	\vdots	\vdots		\vdots
T_n	$CPC(T_n,1)$	$CPC(T_n,2)$...	$CPC(T_n,k)$

那么,多关键词的选择优化就是 M_C 和 M_{CPC} 的一个线性优化问题,即:在约束条件 $\sum_{i=1}^n (M_{cpc}(i,x(i))*M_c(i,x(i))) \leq F$

下,求使 $\sum_{i=1}^n M_c(i,x(i))$ 最大的一组 $x(i)$ 取值。 $x(i)$ 为区间 $[1, n]$ 向区间 $[1, k]$ 的一个映射函数。

由于 M_C 和 M_{CPC} 可推算得出,所以该优化问题必定有解,

估算值取决于您的出价以及地理位置选择。因为点击量估算工具不会考虑您的每日预算,您的广告获得的点击次数可能低于估算值。

图 2 adwords 点击量估算