

投资网络广告关键词的预算分配模型及其简化算法

侯乃聪, 沈向洋

1. 清华大学物理系, 北京 100084
2. 微软亚洲研究院, 北京 100080

[摘要] 探讨了针对网络广告关键词拍卖的预算分配决策的方法和技术, 建立了一个多关键词检索的基本模型和转化模型, 并对模型中的目标函数、约束条件以及生成随机数等进行优化, 得出了在关键词数量有限或相对比较多的两种情况下选择相应解法的结论。

[关键词] 网络经济; 关键词拍卖; 预算约束

[中图分类号] F062.5

[文献标识码] A

[文章编号] 1000-7857(2007)10-0030-04

Adword Auction Model and Its Algorithm

HOU Naicong¹, SHUM Harry²

1. Department of Physics, Tsinghua University, Beijing 100084, China
2. Microsoft Research Asia, Beijing 100080, China

Abstract: A multi-keyword model and its transferred model are built in this paper for the problem of Adword Auction. The algorithm is discussed with respect to the objective function, constraint conditions and the generation of random numbers. Some results are presented in the cases of numerous keywords and a limited number of keywords.

Key Words: internet economics; adword auction; budget constrain

CLC Number: F062.5

Document Code: A

Article ID: 1000-7857(2007)10-0030-04

0 引言

随着 Internet 的发展与普及, 电子商务日益成为企业厂商进行产品和服务营销的一种有效模式, 而网络广告商的出现既为企业厂商与消费者双方带来了极大的便利, 又进一步促进了电子商务的发展。例如, 2004 年, Google 的总收入为 31.89 亿美元, 其中超过 98% 的收入为网络关键词竞拍所得; Yahoo 总收入为 35.74 亿美元, 其中超过 50% 的收入为网络关键词竞拍所得。而两家公司都认为这个市场的潜在价值将有 1 250 亿美元之巨^[1]。

网络广告不同于广播、电视和报刊杂志上的广告, 它通常是由很多不同的商家对某一商品的不同关键词竞价, 这种竞价多次进行并随时更新, 但是并不是所有商家都同时参与每一次竞价。每次竞价更新后, 根据各个关键词的竞价不同, 分别对其由高到低进行重新排

序。排名越高的商家往往享受到更高的关注, 也更容易卖出商品从而获得利润。

由于网络关键词竞价模式的迅速发展及其特别的运行机制, 相关研究大都是分析不同的机制对网络关键词拍卖的影响。如 Benjamin Edelman, Michael Ostrovsky 和 Michael Schwarz^[2]对各种不同的拍卖模式应用和网络关键词竞价机制进行了比较; Hal R. Varian^[3]分析了 Google 和 Yahoo 所用的网络关键词竞拍的运行机制, 并且运用大量的数据从经验上证明了 Google 所运行的机制是合理的; Gagan Aggarwal, Ashish Goel 和 Rajeev Motwani^[4]在假设点击率彼此间不存在独立性的情况下, 建立了 ladder auction 模型; Juan Feng, Hemant K. Bhargava 和 David M. Pennock^[1]则在假设点击率彼此间存在独立性的情况下, 证明了 Ladder Auction 模型与 Adword Auction 模型的一致性。本文从广告商的角度分

收稿日期: 2007-04-20

作者简介: 侯乃聪, 男, 北京市海淀区清华大学物理系, 本科生, 研究方向为网络经济学; E-mail: hac03@mails.tsinghua.edu.cn

沈向洋(通讯作者), 男, 北京市海淀区知春路 49 号微软亚洲研究院, 博士, 研究方向为人机互动、计算机图像;

E-mail: hshum@microsoft.com

析了如何在预算资金的约束下对不同关键词的竞价排序组合进行利益最大化的投资决策。

1 模型构建

1.1 理论假设

文中, 假设在 Internet 上某一商品有 n 个关键词, 并且有 m 个商家参与该商品每个关键词的竞价。此时有一个新的商家 N 想要进入市场, 并且这个商家 N 的广告总预算为 B ; 再假设 $p_{k,i}$ 为这个商家 N 对第 k 个关键词排名为 i 的每一次顾客点击的出价 (这个出价可以保证其排在第 i 名); v_k 为商家 N 对第 k 个关键词的支付意愿; 假设 $c_{k,i}$ 为商家 N 对第 k 个关键词排在第 i 个名次时的最大点击数量,

$$c_{k,i} = \theta_i M_k = \left(\frac{1}{1.428} \right)^{i-1} M_k$$

其中 $\theta_i = (1/1.428)^{i-1}$ 为对第 i 个名次的点击率, M_k 为第 k 个关键词的最大 (第一名) 点击人数。于是预算的分布

是一组数 $x_k, 1 \leq k \leq n$, 并且满足 $\sum_k x_k = B$ 。

1.2 基本模型

令 $U_{k,i}(x)$ 为商家 N 分配到第 k 个关键词投入广告预算 x 的前提下, 对第 k 个关键词出价 $p_{k,i}$ 排在第 i 个名次后所得到的回报, 则有

$$U_{k,i}(x) = \begin{cases} 0 & v_k < p_{k,i} \\ (v_k, p_{k,i}) \frac{x}{p_{k,i}} & v_k \geq p_{k,i}, 0 \leq x \leq c_{k,i} p_{k,i} \\ (v_k, p_{k,i}) c_{k,i} & v_k \geq p_{k,i}, x < 0 \text{ 或 } x > c_{k,i} p_{k,i} \end{cases} \quad (1)$$

设

$$U_k(x) = \max_i U_{k,i}(x) \quad (2)$$

另有

$$\sum_k x_k = B \quad (3)$$

式 (2) 和式 (3) 为模型的约束条件, 同时模型的目标

标为 $\max \sum_k U_k x_k$ 。

1.3 模型转化

由 1.2 节建立的模型可以看出, 对于 $\max \sum_k U_{k,i}(x)$, 确定 x_k 的数值成为解决问题的关键。

另外, 从式 (1) 和式 (2) 中可以看出, 回报函数 $U_k(x)$ 是一个不严格递增的一次函数, 对其求导可得到边际回报函数 u_k , 且 u_k 是一个阶梯函数, 如图 1 所示。假设左边界点为 $a_{k,i}$, 右边界点为 $b_{k,i}$ 。 $a_{k,i}, b_{k,i}, h_{k,i}$ 可以通过式 (1) 得到, 即

$$a_{k,i} = \begin{cases} 0 & i = m-1 \\ \frac{v_k}{p_{k,i+1}} - 1 & i \neq m-1 \end{cases} \quad (4)$$

$$b_{k,i} = c_{k,i} p_{k,i} \quad (5)$$

$$h_{k,i} = \frac{v_k - p_{k,i}}{p_{k,i}} \quad (6)$$

转化后, 模型变为

$$\left. \begin{aligned} \max U(x_0, x_1, x_2, \dots, x_n) &= \sum_{k=1}^n \int_0^{x_k} u_k(x) dx \\ \sum_{k=1}^n x_k &= B \end{aligned} \right\} \quad (7)$$

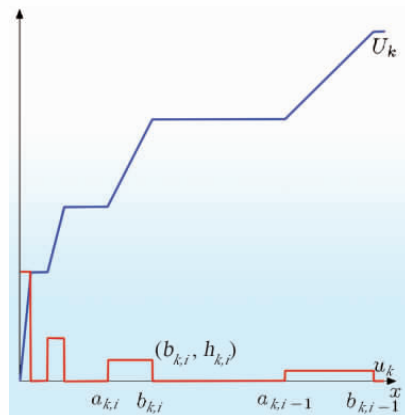


图 1 回报函数及边际回报函数

Fig. 1 Revenue function and marginal revenue function

1.4 模型的求解

对于给定的 n 个关键词、竞拍价格 $p_{k,i}$ 、广告总预算 B 来说, 商家最优的分配总预算的方法最多只有一个关键词的预算分配落在边际回报函数中的右边界点上。据此, 可以利用计算机编程对模型转化后的式 (7) 进行求解。

2 转化模型算法优化

计算机编程求解式 (7) 的计算量是 $O(m^n)$ 。如果关键词的个数比较多, 计算量将是巨大的, 因此需要对算法进行优化。

2.1 目标函数的优化

由于目标函数是一个不连续函数, 故可以将其转化为连续单调 (不严格) 函数。如图 2 所示, 将每一个矩形在面积不变的情况下拉长。由于这只是在 $u_k(x)$ 中的变化, 为了简便起见, 在下面的推导中省略下标 k 。由于

$$a_i = \frac{v}{p_{i+1}} - 1, b_i = c_i p_i$$

则

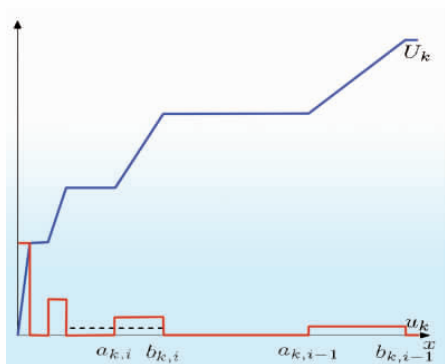


图2 对边际回报函数进行处理

Fig. 2 A scheme dealing with marginal revenue function

$$b_i - a_i = c_i p_i - \frac{p_i(v - p_{i+1})}{p_{i+1}(v - p_i)} c_{i+1} p_{i+1}$$

又因 $S = (b_i - a_i)h_i$, $h_i = \frac{v - p_i}{p_i}$, 则

$$\tilde{h}_i = \frac{S}{b_i - a_i} = \frac{c_i v - c_i p_i - c_{i+1} v + c_{i+1} p_{i+1}}{c_i p_i - c_{i+1} p_{i+1}}$$

下面证明 $\tilde{h}_i$ 是单调的。

$$\tilde{h}_i - \tilde{h}_{i-1} = \frac{c_i v - c_i p_i - c_{i+1} v + c_{i+1} p_{i+1}}{c_i p_i - c_{i+1} p_{i+1}} - \frac{c_{i-1} v - c_{i-1} p_{i-1} - c_i v + c_i p_i}{c_{i-1} p_{i-1} - c_i p_i} \quad (8)$$

由于 $c_{i-1} > c_i > c_{i+1}$, $p_{i-1} > p_i > p_{i+1}$, 式(8)通分后, 分子为 $0.428 \times 1.428 c_{i+1}^2 v [1.428(p_{i-1} - p_i) - (p_i - p_{i+1})]$ 。由于模型的背景是多次竞拍, 故假定经过充分多次的竞拍后, 关于排名的竞价是稠密的, 即任何两个相邻名次间的竞价差非常小, 甚至相等。所以有

$$0.428 \times 1.428 c_{i+1}^2 v [1.428(p_{i-1} - p_i) - (p_i - p_{i+1})] \geq 0$$

故 $\tilde{h}_i - \tilde{h}_{i-1} \geq 0$, 即 $\tilde{h}_i$ 是单调的, 如图3所示。对该函数的奇点进行光滑处理后可保证函数 u'_k 为一单调(不严格)可导函数。

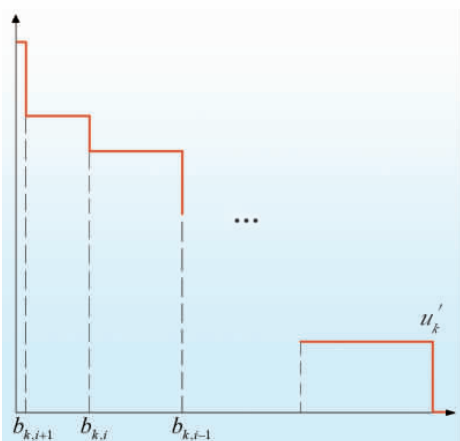


图3 处理后的边际回报函数

Fig. 3 Revised marginal revenue function

由此, 可对函数 $U(x_0, x_1, x_2, \dots, x_n) = \sum_{k=1}^n \int_0^{x_k} u'_k(x) dx$ 求偏导数。

即当 $\frac{\partial U}{\partial x_0} = \frac{\partial U}{\partial x_1} = \frac{\partial U}{\partial x_2} = \dots = \frac{\partial U}{\partial x_n} = 0$ 时,

$$u'_1(x_1) = u'_2(x_2) = \dots = u'_{n-1}(x_{n-1}) = u'_n \left(B - \sum_{i=1}^n x_i \right)$$

由于 u'_k 为单调函数, 故可得唯一一组 $(x_0, x_1, \dots, x_n)$ 使 U 取到最大值。

将 u'_k 还原为 u_k , 并令 $x_k (1 \leq k \leq n)$ 向左侧移动到最精确的右边界点处 (已经处在右侧边界点的则不用移动), 得到一组新点 $(x'_1, x'_2, \dots, x'_n)$, 设 $C' = B - \sum_{k=1}^n x'_k$, 当

$C' = 0$ 时, 即为所求。当 $C' > 0$ 时, 则可在 $(x_0, x_1, x_2, \dots, x_n)$ 的基础上用 C' 作为约束条件求得最大值, 由于 C' 相对于 B 来说并不大, 可直接用穷举法。

但这种方法存在一定的局限性, 因为局部最优并不代表整体最优。所以在 $C' > 0$ 时所得到的理论最优解与实际最优解存在差距, 并且这个差距的上界难以估算。

2.2 增加约束条件

模型中, n 个关键词中最多只有一个关键词的预算不落在函数 u_k 的右侧边界点上。假设每个关键词的预算均落在函数 u_k 的右侧边界点上, 那么在原约束 $\sum_{k=1}^n x_k = B$ 条件下可求得最优解, 并且如果剩余的预算不够任意一个关键词的预算时, 这部分剩余的预算将右移到下一个右侧边界点处, 故此时与实际最优解的误差上限是可以求得的。这样, 就可以将式(7)转化为

$$\left. \begin{aligned} \max Z &= \sum_{i=1}^n \sum_{j=1}^m r_{i,j} x_{i,j} \\ \sum_{i=1}^n \sum_{j=1}^m C_{i,j} x_{i,j} &\leq M \\ \sum_{j=1}^m x_{i,j} &\leq 1 \\ x_{i,j} &= \begin{cases} 0 & \text{在一定的约束下不执行} \\ 1 & \text{在一定的约束下执行} \end{cases} \end{aligned} \right\} \quad (9)$$

式中, $i=1, 2, \dots, n; j=1, 2, \dots, m; r_{i,j}$ 为收益, 且 $r_{i,j} = \sum_{k=1}^j (b_{i,k} - a_{i,k}) h_{i,k}$; $x_{i,j}$ 为决策变量, 表示是否应该执行的方案; $C_{i,j}$ 为方案的初始成本; M 为资金预算约束; n 为行的数目 (独立方案的数目); m 为列的数目 (相互排斥方案的数目)。

此时, 预算投资决策问题转化为整数规划中的 0-1 规划的问题, 可用运筹学应用软件直接求解。另外, 也可以用分量 $x_{i,j}=0$ 或 1 来分解问题, 这个算法 (不一定

是相关软件中所采用的算法)称为 0-1 规划中的隐枚举法,其中运用了分支定界法的思想^[5],但无法得到理论最大值与实际最大值之间的精确差距,而且不能估算出这个差距的上界。

2.3 生成随机数

在原约束条件 $\sum_{k=1}^n x_k = B$ 下,可以生成 m 组,每组 n 个随机数。把这 n 个随机数带入原目标函数 $U(x_0, x_1, x_2, \dots, x_n) = \sum_{k=1}^n \int_0^{x_k} u_k(x) dx$ 可得一具体数值。进行 m 次操作后,选择最大值作为实际最优解的一个近似。该方法的缺点是无法得到理论上的近似最大值与实际最大值之间的差距,并且不能估算出差距的上界。

3 结论

本文构建了投资网络关键词的预算分配模型(式(7))并进行简化。在关键词不多的情况下,可用穷举法解决关键词竞价分配预算的决策问题,而关键词较多时,如果对结果的精度要求不高,可以考虑优化目标函

数、增加约束条件和生成随机数等 3 种方法。

参考文献 (References)

- [1] FENG Juan, BHARGAVA H K, PENNOCK D M. Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms [J]. Inform Journal on Computing, 2007, 19 (1): 137-148.
 - [2] EDELMAN B, OSTROVSKY M, SCHWARZ M. Internet advertising and the generalized second price auction[R]. US: National Bureau of Economic Research, 2005: 11765.
 - [3] VARIAN H R. Position auctions [R]. Berkeley: University of California, 2005.
 - [4] AGGARWAL G, GOEL A, MOTWANI R. Truthful auctions for pricing search keywords [C]// Proceedings of ACM Electronic Commerce, 2006.
 - [5] 吴祈宗. 运筹学最优化方法 [M]. 北京: 机械工业出版社, 2005: 197-202.
- WU Qizong. Operational research and methods of optimization [M]. Beijing: China Machine Press, 2005: 197-202.

(责任编辑 代丽)

·科技智力游戏·

如果您在 6 月 15 日前将本栏目答案连同期刊所附调查表寄至本刊编辑部,您将有获得自下期至 2008 年底的《科技导报》一套。

轻松一刻——九宫填数

该游戏的填数规则为:①在 9×9 的大九宫格内,出题者已给出若干数字,其他宫位的数字空留,需要读者根据逻辑推理填写剩下空格里的数字;②填满数字的大九宫里,数字之间的关系必须满足:每一行与每一列都有 1~9 的数字,每个实线所标 3×3 的小九宫格里也有 1~9 的数字,且每个数字在每行、每列及每个小九宫格里只能出现一次,即既不能重复也不能缺少;同时在 9×9 大九宫格的两条对角线上,均有 1~9 的数字,且只能出现一次。答案请见下期。

九宫1 难度系数 ◆◆◆◆◇◇

			6	2				8
2		9						
							4	
	3	5			6			9
6			7		5			4
7		2			3	1		
	5							
						1		6
	4			8	1			

九宫2 难度系数 ◆◆◆◆◇◇

1			2					4
			9	5	6			
		3		4		9		
	9	6	3		2	1	5	
		7		8		4		
			4	3	1			
7				9				3

上期答案

九宫1 难度系数 ◆◆◆◆◇◇

1	3	9	6	8	7	5	4	2
8	7	5	2	1	4	9	6	3
2	6	4	9	3	5	7	8	1
6	9	2	3	4	1	8	7	5
7	5	1	8	9	2	4	3	6
4	8	3	5	7	6	1	2	9
5	1	8	7	6	3	2	9	4
9	4	6	1	2	8	3	5	7
3	2	7	4	5	9	6	1	8

九宫2 难度系数 ◆◆◆◆◇◇

2	9	3	7	5	1	6	8	4
6	1	8	3	9	4	2	7	5
4	5	7	6	8	2	1	9	3
8	7	9	4	1	3	5	6	2
3	2	5	8	6	7	9	4	1
1	6	4	5	2	9	8	3	7
5	4	2	9	7	8	3	1	6
7	8	1	2	3	6	4	5	9
9	3	6	1	4	5	7	2	8

(供稿 李建才 责任编辑 赵佳)