



# 南京大學

## 本科畢業論文

院 系 計算機科學與技術系

專 業 計算機科學與技術

題 目 第一價點擊量拍賣模型研究

年 級 2010 學 號 101220153

學生姓名 曾明陟

指導老師 吳駿 謝俊元 職 稱

論文提交日期 2014 年 6 月

# 南 京 大 学

## 本科生毕业论文（设计、作品）指导情况记录

开 题 简 况	<p>题目：</p> <p>1、选题质量（简述选题与专业培养目标、专业要求关系、题目难度、工作量、创新性、理论性、实用性）</p> <p>2、开题意见：</p> <p>指导教师签名：</p> <p>年 月 日</p>
中 期 检 查	<p>指导教师检查论文的进展情况：（指导和培养学生查阅文献资料、综合运用知识、研究方案设计、研究方法和手段运用和外文应用等能力简况）</p> <p>指导教师签名：</p> <p>年 月 日</p>

# 南 京 大 学

## 本科生毕业论文（设计、作品）指导教师评阅意见

指导教师评语：

指导教师签名：

年 月 日

# 南 京 大 学

## 本科生毕业论文（设计、作品）评阅教师评阅意见

评阅教师评语：

评阅教师签名：

年 月 日

# 南 京 大 学

## 本科生毕业论文（设计、作品）答辩记录、成绩评定

答辩记录：

答辩记录人签名：

答辩小组评语：

答辩小组成员：\_\_\_\_\_

成绩\_\_\_\_\_

组长签名：

答辩时间： 年 月 日

# 南京大学本科生毕业论文（设计、作品）中文摘要

毕业论文题目： 第一价点击量拍卖模型研究

计算机科学与技术 院系 计算机科学与技术 专业 2010 级

本科生姓名： 曾明陟

指导教师（姓名、职称）： 吴骏 谢俊元

摘要：

搜索引擎正在逐渐成为人与海量信息、智能信息交互的统一入口，而付费搜索则是支撑其庞大研发支出的重要支柱。传统的付费搜索模型对固定的广告位进行拍卖，难以满足用户按需定制的需求。本文结合点击量拍卖的概念以及比例分配模型，提出了一种第一价点击量拍卖 FCA (First-price Click Auction, FCA) 模型。FCA 模型通过动态分配算法将固定的广告位转化为预期点击量，并通过比例分配式的竞价规则对预期点击量进行拍卖，使得广告主可以根据自身预算设计适宜的竞价策略，获得相匹配的点击量。我们从理论上分析了模型的性质以及买家的竞价策略，证明了模型存在唯一的纳什均衡。随后我们将 VCG 机制作为基线机制，利用来自 Yahoo! Labs 的实际竞价数据集进行了数值模拟。通过实验我们发现，FCA 模型不仅使用户得到与自身出价更加匹配的投资回报，同时也使得点击量得到更加有效的分配。与此同时，FCA 模型在盈利能力上则略逊 VCG 机制。

关键词：搜索引擎，付费搜索，关键词拍卖，点击量拍卖，比例分配

# 南京大学本科生毕业论文（设计、作品）英文摘要

THESIS: Click Auction Model based on Proportional-Share Allocation

DEPARTMENT: Department of Computer Science and Technology

SPECIALIZATION: Computer Science and Technology

UNDERGRADUATE: Mingzhi Zeng

MENTOR: Jun Wu and Junyuan Xie

## ABSTRACT:

Search engines are turning into the ultimate form of retrieving and interacting with big and intelligent data. In the near future, online search will become the unified interface between human and massive information. Sponsored search, which is the fundamental way of profiting of search engines, is not flexible enough for advertisers since existing mechanisms sell slots with fixed expected clicks. Advertiser who has his unique budget and objective will want to get clicks tailored to his need. In this paper we introduce a new model named First-price Click Auction Model based on Proportional-Share Allocation (or First-price Click Auction Model for short). The FCA model introduced several algorithms to convert fixed slots into expected clicks, and an auction to allocate clicks proportionally according to bidders' bids. We also present theoretical analysis on players' best response, proof of the existence and uniqueness of Nash Equilibrium, and finally experiment result on real-world dataset from Yahoo! Labs. We discovered that, based on our experiment, the FCA model allocates better matched share of clicks to advertisers. Also, the FCA model has an overall performance in efficiency. Meanwhile, the new model's platform revenue is slightly outperformed by the VCG mechanism.

KEY WORDS: Search Engines, Sponsored Search, Paid Search, Keywords Auction, Click Auction, Proportional-Share Allocation

# 目录

1 绪论 .....	1
1.1 研究背景 .....	1
1.1.1 在线广告与付费搜索 .....	1
1.1.2 关键词拍卖 .....	3
1.2 本文工作 .....	5
1.3 本文结构 .....	5
2 关键词拍卖模型 .....	6
2.1 运作流程 .....	6
2.2 模型定义 .....	6
2.3 拍卖规则 .....	8
2.4 现有拍卖机制 .....	9
2.4.1 GFP 拍卖机制 .....	9
2.4.2 GSP 拍卖机制 .....	10
2.4.3 VCG 拍卖机制 .....	11
2.4.4 阶梯式拍卖机制 .....	13
2.5 保留价 .....	14
3 第一价点击量拍卖(FCA)模型 .....	14
3.1 模型假设 .....	15
3.2 出价规则 .....	16
3.3 排序规则 .....	16



3.4 支付规则 .....	19
3.5 分配规则 .....	19
3.7 纳什均衡存在性与唯一性.....	21
4 数值模拟 .....	22
4.1 数据来源 .....	22
4.2 数据预处理 .....	22
4.3 模拟方法与评估标准.....	23
4.4 总体分析 .....	23
4.5 有效竞价人数对 FCA 模型的影响.....	27
4.6 点击量分布对 FCA 模型的影响.....	28
4.7 模型迭代次数 .....	29
4.8 广告位重叠 .....	29
4.9 极端情况分析 .....	30
5 结论 .....	31
参考文献.....	33
致谢.....	35

## 图表

图 1.1	网络付费搜索的参与者、目标与过程.....	3
图 1.2	锯齿模式 .....	4
图 4.1	CFCA 与 VCG 效率比值分布 .....	24
图 4.2	CFCA 与 VCG 平台收益比值分布 .....	25
图 4.3	广告主点击量份额与支付相关系数分布.....	26
图 4.4	a) 广告位个数对模型效率比分布的影响.....	27
图 4.4	b) 广告位个数对平台收益比分布的影响.....	27
图 4.5	a) 点击量标准差对模型效率比分布的影响.....	28
图 4.5	b) 点击量标准差对平台收益比分布的影响.....	28
图 4.6	CFCA 达到均衡前的迭代次数分布 .....	29
图 4.7	CFCA 交易中最大覆盖广告位分布 .....	30
表 4.1	Yahoo! Labs 数据集表项.....	22
表 4.2	CFCA 与 VCG 对比总览 .....	23
表 4.3	CFCA 模型与 VCG 效率比值总览 .....	25
表 4.4	CFCA 模型与 VCG 平台收益比值分布 .....	25
表 4.5	极端案例下的交易细节.....	31

## 1 绪论

付费搜索(Sponsored Search)是目前搜索引擎在线广告的主要形式。在付费搜索中,广告主根据自身营销策略对某个关键词进行拍卖,搜索引擎根据全部广告主的出价对广告进行相应的排序,并在用户检索时显示在相应的位置。在关键词被点击后,搜索引擎对广告主收取相应的费用。

付费搜索机制在搜索引擎的发展中扮演着重要的作用,不仅为提供商创造了丰厚的利润,缔造了 Google、Yahoo! 等巨擘,同时也创造了一种前所未有的商业模式,使得来自全球各地的商家能够以极低的门槛推广自己的产品。从用户的角度来看,付费搜索也使得他们能够更方便快捷地找到适合自己的商品与服务。随着付费搜索的快速发展,人们开始关注以下问题:1. 如何使得广告位得到最有效的分配? 2. 如何使搜索引擎获得更高的利润? 由于关键词拍卖是付费搜索的主体,对这些问题的研究也随之转化为对关键词拍卖机制的研究。本文中我们将探讨现有机制在广告位分配方式上的不足,并提出一种新的拍卖模型—第一价点击量拍卖(First-price Click Auction, FCA)模型,来改善点击量分配。

### 1.1 研究背景

#### 1.1.1 在线广告与付费搜索

自 1994 年 Hot Wired 推出首个可点击旗帜广告起,在线广告已经经历了近 20 年的发展。随着技术的进步,在线广告已从几乎一片空白发展为品种多样化并且定制化的庞大产业。据 IAB 统计[5],2013 年美国在线广告销售达到近 428 亿美元,首次超过电视广告销售额,全球在线广告的投入则达到 1042 亿美元。这一数字还在以惊人的速度增长中。互联网广告为广大的互联网公司提供了维护和研发所需的大量资金,可以说没有在线广告就没有今天的互联网。

最初的在线广告形式以旗帜广告为主,广告提供商将网站的某个区域作为广告区域,并向广告主开放,通常由广告主和提供商进行协商以确定某个席位的价

格，以 Hot Wired 作为这类广告形式的代表。随后从 Netscape 和 Infoseek 开始，在线广告逐渐将计费模式改为 CPM(Cost Per Mille，每千人成本)，但基本的广告形式仍然采用旗帜广告的形式。在此之后陆续出现了弹窗广告、悬浮广告、邮件广告等诸多在线广告形式，与此同时自动竞价排名机制也被引入广告计费机制当中。

1997 年，Overture 的原型 Goto.com 推出了关键词拍卖的付费搜索计价方式，从此拉开付费搜索的序幕。2000 年，成立两年的 Google 公司首次推出了 AdWords 服务，从 CPM 收费模式开始转向按照点击量进行收费(Cost per Click，即每次点击成本)的付费搜索模式。由于该付费搜索模式的低门槛高灵活性的特点，AdWords 开始迅速被人们所接受，并使 Google 公司逐渐成为搜索领域的领头羊。

随着付费搜索的发展，如何更好地满足用户和广告主的需求就成为了研究者们关注的议题。为对此进行更加细致的分析，Jansen 等[2]提出了如下的基本要素：

- *提供者*：需要为某个站点导向流量的个人或组织
- *提供者内容*：一组关键词以及相关的 URL 与描述。这些内容常常带有强烈的商业意味。
- *搜索引擎*：接受用户请求并返回相关的网页或内容
- *搜索引擎审查过程*：搜索引擎用来确保提供者内容与关键词相关的一套方法
- *搜索引擎关键词与内容索引*：一种将提供者的关键词与用户请求关联起来的机制
- *搜索引擎用户接口*：用于将提供者内容按照某种方式排序后呈现给用户的应用
- *搜索引擎追踪*：将关键词与请求相匹配并收集提供着内容、出价，并向提供者收费的方式
- *搜索者*：在搜索过程中点击某个付费搜索链接的用户

基于上述要素得出的付费搜索完整模型如图 1.1 所示[2]。

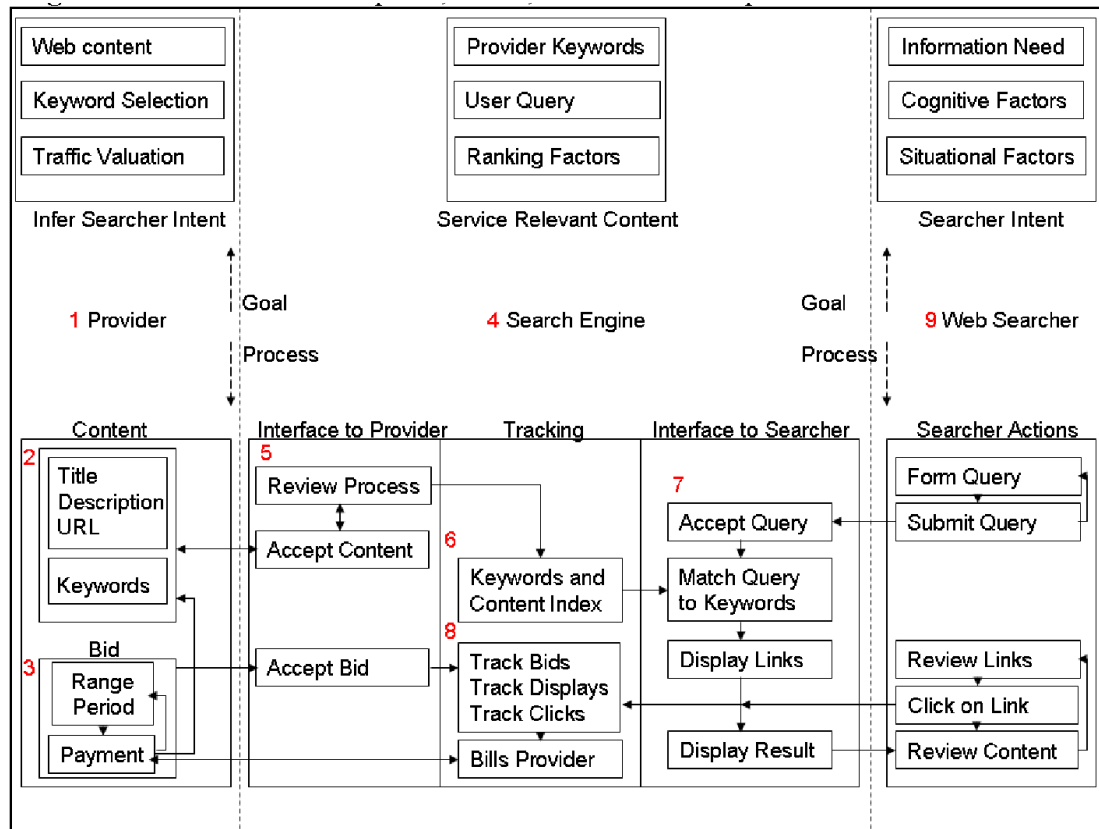


Fig. 1.1 The participants, goals, and process of sponsored search on the web

图 1.1 网络付费搜索的参与者、目标与过程

### 1.1.2 关键词拍卖

在整个搜索流程中,对于搜索引擎而言最重要的是对搜索结果进行排序的部分。排序算法一方面需要决定哪些来自网络的结果应该排在最前面,以使用户最快地找到所需信息;另一方面还需要决定哪些广告主的广告应该显示在最醒目的位置,以使的搜索引擎获得最大的利润。目前的绝大多数搜索引擎仍然通过 Overture 提出的关键词拍卖的方式来决定广告位的排位。在这种拍卖模型中,广告主不再对某一个公共的广告位进行出价,而是对于某一个感兴趣的关键词给出一个报价,当用户搜索这一关键词时,搜索引擎将会根据广告主的报价从高到低排序,并将其广告显示在相应的位置。同时,每当用户点击一次广告,搜索引擎便会按照广告主的报价收取相应的费用。

由于关键词拍卖从形式上与普通拍卖并无差别,因此早期的关键词拍卖机制借鉴了许多已有的拍卖机制。最常见的一种便是由英式拍卖衍生而来的广义第一

价拍卖。这种拍卖中买家可以连续出价，在最终结算时以最后一次出价为准，买家被按照报价高低排序，最终支付的费用即为其最终的出价。

这种机制不久后被证明具有一定的局限性。首先，能够第一时间对对手的出价进行反馈的买家将会占据优势，这将会导致一些无意义的出价产生。其次，这样的竞价方式将会导致买家进行无止尽的重复出价，这个著名的缺陷被称为“竞价战争(bidding war)”或“锯齿模式”[1]。

**例 1.1**：假设有两个买家竞争排序最高的广告位，买家 1 和买家 2 对该广告位的估价分别为 12 美元和 10 美元。为了获得该广告位，买家 1 和买家 2 将会不断在对方出价的基础上提高 0.1 美元。当买家 1 出价 10.1 美元时，买家 2 已经无利可图，于是将自己的出价调整到最低出价 6 美元。此时买家 1 不再需要出价 10.1 美元来获得排序最高的广告位，于是将自己的出价降到 6.01 美元。这样两个买家的出价不断循环，形成竞价战争。

在早期的第一价拍卖中，锯齿模式的确存在，如图 1.2 中就展示了 Overture 引擎中的某关键词在 2002 年 6 月中的竞价所呈现的“锯齿”模式[1]。

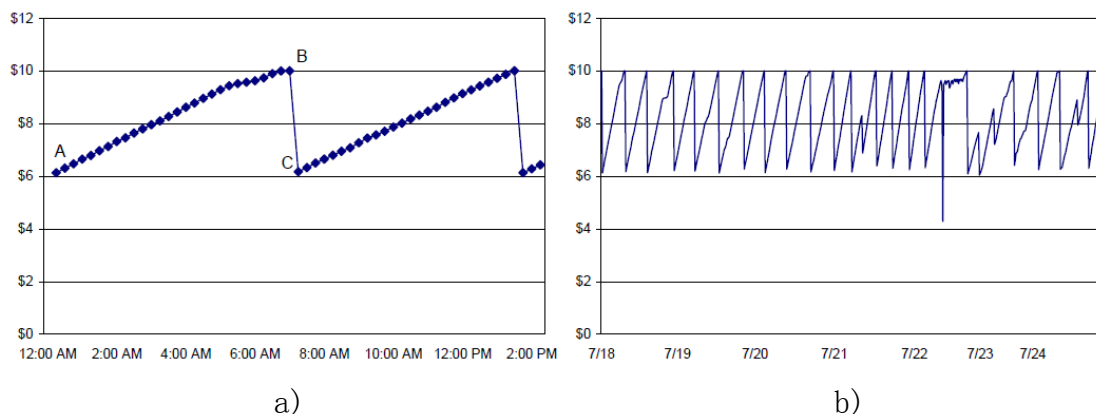


Fig. 1.2 “Sawtooth” Pattern

图 1.2 锯齿模式

事实上，这种早期机制中不存在均衡，用户的行为是难以预测的。这对于搜索引擎的盈利是不利的。

从搜索引擎的角度来看，最理想的情况便是存在唯一的均衡，并且均衡中每个广告主的报价均为其真实的估价。这样的机制被称为 Truthful 机制。目前部分搜索引擎采用的 VCG 机制[16]便实现了这一点。

现实中关键词机拍卖制则不一定以 Truthful 为目标，例如 Google 率先开始采用的 GSP 机制[1]。Edelman 等证明了 GSP 机制中的买家没有占优策略，但 GSP

机制的收益却高于 VCG 机制。

关键词拍卖机制的设计需要考虑的因素非常多,研究者们针对不同的需求提出了不同的改良机制。例如 Hafalir[21],Feldman[18]等提出在广告主的预算受到限制的情况下的最优拍卖;Aggarwal 等[4]讨论了加权 VCG 在某些条件下不适用的缺陷,并提出了加权降序排序下最优的拍卖机制;Hu[19]则讨论了保留价对于第一价和第二价拍卖的影响以及最优保留价的设置问题。我们将在第 2 章详细介绍部分重要的关键词拍卖机制,并对其优势与劣势进行分析。

## 1.2 本文工作

现有的关键词拍卖绝大多数以广告位为单位进行拍卖,这样的拍卖方式不利于买家获取与自身需求相符的点击量。我们提出一种第一价点击量拍卖机制(简称为第一价点击量拍卖,First-price Click Auction),将固定的广告位转化为预期点击量,并以预期点击量为单位进行拍卖。我们从理论上分析了机制中买家的最适反应,并证明了机制存在唯一的纳什均衡。最后我们通过大量数据对 FCA 模型和 VCG 模型进行模拟,并对结果进行了对比分析。我们的结果显示 FCA 模型在减少极少的系统收益的情况下,实现了更好的点击量分配能力,同时对系统效率也有一定程度的提高。

## 1.3 本文结构

本文首先介绍付费搜索的发展历史以及相关研究,随后介绍 FCA 模型,最后通过数值实验将 FCA 模型与 VCG 模型进行对比。全文按照以下顺序组织:第一章,介绍付费搜索的发展历史;第二章,介绍关键词拍卖的模型与规则,介绍现有的 GSP、VCG 等机制及其性质,并介绍一些相关的研究方向;第三章,介绍 FCA 模型的规则与理论分析,讨论了买家的最适反应,并证明了 FCA 模型存在唯一的纳什均衡;第四章,通过数值模拟将 FCA 模型与 VCG 模型进行对比,分析模型的总体表现以及各个因素对于模型性能的影响;第五章进行总结与展望。

## 2 关键词拍卖模型

第 1 章中我们介绍了关键词拍卖在付费搜索中的作用，并简略地介绍了几个关键词拍卖机制。在本章中我们系统地对关键词拍卖模型进行定义，并对几个目前常见的关键词拍卖机制进行详细分析。

### 2.1 运作流程

1. 广告主（买家）提交广告，以及与广告相关的描述、链接等；
2. 广告主为广告选取若干个关键词，并对这些关键词给出单次点击的报价；
3. 搜索引擎收集并整理某一个关键词的所有买家出价，按照一定的规则对相应的广告进行排序；
4. 当用户搜索某一个关键词时，搜索引擎检索得到该关键词对应的广告与排名，将广告展示在相应的位置上，并根据支付规则收取广告主相应的费用。

关键词拍卖是连续的拍卖，也就是说 1 和 2 可以不断地进行，广告主可以在任何时刻提交或更改对于某个关键词的报价。而只有在 4 发生时，一次结算才会进行，此时用于结算的报价为每个广告主在此之前的最后一次报价。

### 2.2 模型定义

对于一个给定的关键词，在某一时刻  $t$  有  $N$  个广告位和  $K$  个广告主。第  $i$  个广告位一段时间内收到的点击量为  $\alpha_i$ ，对于广告主  $k$ ，每次点击带来的收益为  $v_k$ ，广告主  $k$  在时刻  $t$  之前的最后一次出价为  $b_k$ ，若有多个广告主出价相同，则将他们随机排序。若在  $t$  时刻有用户检索该关键词，则搜索引擎对所有广告主的出价进行排序，用  $b^j$ 、 $p^j$  和  $g^j$  分别表示第  $j$  高的出价最终支付的价格和对应的广告主。所有的广告主均为风险中立，广告主  $k$  在第  $i$  个广告位的收益为  $\alpha_i v_k$  减去搜索引擎收取的费用。注意这里的定义假设广告主在不同位置收到的点击量均是相同的，广告主对于不同广告位中每次点击的估价也是相同的。



我们将涉及到的符号罗列如下，方便参考：

- $g$  - 表示买家
- $b$  - 表示买家的出价
- $w$  - 表示买家的权重
- $p$  - 表示买家最终支付的价格向量
- $v$  - 表示买家对单次点击的估价
- $\alpha$  - 表示广告位的点击量向量
- $c$  - 表示买家最终获得的预期点击量
- $\mathcal{S}$  - 表示广告位的点击量之和
- $E$  - 表示机制效率
- $\mathcal{R}$  - 表示平台收益
- $u$  - 表示买家的效用函数

此外我们对文中的一些概念进行定义：

**定义 2.1**：点击量(Click-through)

一段时间内广告位被点击的次数。

**定义 2.2**：点通率(Click-through Rate)

点通率表示一个广告位在一段时间内的点击量与浏览量的比，通常用于评估广告的质量。

**定义 2.3[4]**：可分点通率(Separable Click-through Rates)

用 $CTR_{i,j}$ 表示买家  $i$  被放置在第  $j$  个广告位时的点通率。如果存在 $\mu_1, \mu_2, \dots, \mu_n > 0$ 和 $\theta_1 \geq \theta_2 \geq \dots \geq \theta_K > 0$ ，使得 $CTR_{i,j} = \mu_i \theta_j$ ，则点通率是可分的。

**定义 2.4[17]**：最适反应(Best Response)

在一个博弈中，一名参与者的最适反应是在其他参与者策略给定时，他所选择的

将自己的收益最大化的策略或策略组合。

**定义 2.5[15]：** 纳什均衡点

一个  $n$  元策略组  $\mathcal{K}$  是一个纳什均衡点，当且仅当对所有的  $i$  都有：

$$p_i(\mathcal{K}) = \max_{r_i} p_i(\mathcal{K}; r_i)$$

其中  $\mathcal{K} = (s_1, \dots, s_n)$ ，而  $(\mathcal{K}; r_i)$  则表示  $(k_1, \dots, k_{i-1}, r_i, k_{i+1}, \dots, k_n)$ 。当博弈处于均衡点时，每个参与者的策略都是其最适反应。因此任何参与者都不能通过改变自己策略的方式来提高自己的收益。

## 2.3 拍卖规则

拍卖规则可以大致划分为出价规则、排序规则和支付规则。

**出价规则：**对于搜索引擎而言，用户的出价是判断用户类型的重要依据，因此搜索引擎总是希望用户的出价简单明了。通常付费搜索拍卖采用的出价规则均为连续出价，即用户可以在任意时间修改其出价。在最终进行排序和计费时，均按照最后一次出价作为用户的出价。

**排序规则：**通常采用的排序规则有“ $v$  ranking”和“ $v \times \alpha$  ranking”两种，前者为早期的 Overture 采用，后者为 Google 的第二价拍卖采用。 $v$  代表广告主的出价，这里的  $\alpha$  代表广告与关键词的相关性。因此，“ $v$  ranking”方式中，广告主投放的广告仅按照广告主为该关键词的出价从高到低排序，而“ $v \times \alpha$  ranking”方式中，广告主投放的广告按照出价与相关性的乘积进行排序。在实际应用中，“ $v \times \alpha$  ranking”在几乎所有的案例中的表现都要好于其他规则[3]。

**支付规则：**早期的搜索引擎采用的均为第一价支付，用户给出的报价即为最终收取的费用。这种支付规则在前文中已经证明存在缺陷。因此 Google 在后来的 AdWords 中采用了通用第二价的竞价机制，这种机制中，排名第  $i$  的买家只需支付排名第  $i+1$  位的买家的出价。另一种支付规则由 VCG 机制定义，这种机制中每个用户需要支付的费用为其后所有用户的负外部性之和。

## 2.4 现有拍卖机制

### 2.4.1 GFP 拍卖机制

广义第一价拍卖 (Generalized First Price, GFP)，是最为直观的一种关键词拍卖方式。拍卖方将所有买家的出价按照递减的顺序排列，然后依次为出价第  $i$  高的买家分配点击量第  $i$  高的广告位，每个获得广告位的买家最终支付的价格即为他的出价。在 2004 年以前 Yahoo 公司一直采用这种拍卖机制来销售其在线广告。

值得注意的是，这种机制不同于普通的单物品第一价拍卖机制，它在一般情况下并没有纯策略纳什均衡。

**定理 2.1**：在 GFP 中只有当  $\alpha_1 = \dots = \alpha_{\min(N,K)}$  时才存在纯策略纳什均衡。

**证明**：设  $\mathbf{b} = (b^1, \dots, b^{\min(N,K)})$  为一个纯策略纳什均衡，首先可以证明  $b^i = b^{i+1}$ ，如果  $b^i > b^{i+1}$ ，那么买家  $i$  总可以更改自己的出价为  $b^i - \varepsilon > b^{i+1}$ ， $\varepsilon$  可以任意小，此时买家  $i$  仍然排在第  $i$  位而收益可以增加  $\alpha_i \varepsilon$ ，因此  $\mathbf{b} = (b^{\min(N,K)}, \dots, b^{\min(N,K)})$ 。其次可证明  $b_{\min(N,K)} = 0$ ，否则买家  $n$  可以通过调整其出价为  $b_{\min(N,K)} - \varepsilon$ ，从而仍然排在最后一位但收益增加  $\alpha_{\min(N,K)} \varepsilon$ ，因此  $\mathbf{b} = (0, \dots, 0)$ 。进一步分析，对于任意买家  $j$ ，如果存在改变报价的动机，设改变后的报价为  $\varepsilon > 0$ ，则必须满足  $(v_j - 0)\alpha_j < (v_j - \varepsilon)\alpha_1$ ，也就是满足  $0 < \varepsilon < v_j \frac{\alpha_1 - \alpha_j}{\alpha_j}$ 。

因此如果要使纯策略纳什均衡成立，则必须有  $\alpha_1 = \alpha_j$ ， $\forall j \in 1 \dots \min(N, K)$ ，即

$\alpha_1 = \dots = \alpha_{\min(N,K)}$ 。

### 2.4.2 GSP 拍卖机制

广义第二价拍卖(Generalized Second Price, GSP) [1], 最早由 Google 公司开始采用, 并大获成功。这种机制与 GFP 不同的地方在于其支付方式, 排名第  $i$  位的买家只需支付第  $i+1$  高的出价即可获得第  $i$  个广告位。

**定理 2.2[1]** : GSP 不是一个 truthful 机制

**证明:** 只需举一个反例即可。假设有 3 个买家和 2 个广告位, 买家的估价分别为: 10 美元、4 美元和 2 美元, 广告位在一段时间内的点击量分别为 200 和 199。如果所有买家都真实汇报自己的价格, 那么买家 1 的收益为:  $(10-4)*200=1200$  美元。如果买家 1 不汇报真实出价, 而是只给出 3 美元的出价, 那么他的收益将变为  $(10-2)*199=1592$ , 大于 1200。

### 静态博弈与局部无嫉妒均衡

由于 GSP 采用连续出价的规则, 我们可以将其视为一个无限重复的博弈。在这个博弈当中, 参与者原本只有本身的信息, 但在不断重复博弈的过程当中他们可以获得其他参与者的类型, 并据此改变自己的策略。理论上讲, 如果参与者能够获得足够多的信息, 他们就可以设计自动竞价程序来不断修改自己的报价以获得最大的收益。但现实中搜索引擎并不允许买家使用未经授权的自动程序进行竞价, 因此我们只讨论买家可能执行的简单策略。我们有以下假设:

1. 随着竞价次数的增长, 每名买家都可以获知所有买家的报价
2. 静态均衡中的出价必须是所有买家对其他买家的最适反应(Best Response)

如果某名买家想要提高自己的排名, 首先需要考虑的就是如何占有自己身前一位的买家的广告位。如果排名  $i+1$  的买家提高自己的出价, 那么在 GSP 中买家  $i+1$  的收益并不会减少而买家  $i$  的收益将会减少。此时买家  $i$  也可以通过降低自

己的出价至刚好低于买家  $i+1$  的出价来“报复”。当双方最终到达一个稳定的状态，互相没有动机改变报价时，我们称为局部无嫉妒均衡。

**定义 2.6[1]：** 局部无嫉妒均衡

在 GSP 中，如果一个静态博弈的均衡中，任意一个参与者都不能通过超越前一个参与者来提高自己的收益，则称此均衡为局部无嫉妒均衡。在局部无嫉妒均衡中，对于任意的  $i \leq \min(N, K)$ ，都有  $\alpha_i v_{g^i} - p^i \geq \alpha_{i-1} v_{g^i} - p^{i-1}$

随着拍卖的进行，买家会不断地改变自己的出价，如果买家的出价最终收敛到一个静态的向量，那么这个静态平衡一定满足静态博弈  $\Gamma$  的局部无嫉妒均衡。[1]中将局部无嫉妒均衡的集合映射到双边市场匹配问题中的稳定分配的集合，并证明了如下定理，我们不加证明直接给出：

**定理 2.3[1]：** GSP 的静态博弈  $\Gamma$  的局部无嫉妒均衡中的任何结果都是一个稳定分配

**定理 2.4[1]：** 如果买家的数目超过物品的数目，那么任何稳定分配都是静态博弈  $\Gamma$  的局部无嫉妒均衡

**定理 2.5[1]：** 策略组合  $B^*$  是静态博弈  $\Gamma$  的一个局部无嫉妒均衡。这个均衡中任意一个参与者的排名和支付金额与 VCG 机制的优势策略均衡中相同。 $\Gamma$  中其他局部无嫉妒均衡的卖家总收益都不小于  $B^*$ 。

### 2.4.3 VCG 拍卖机制

VCG(Vickrey-Clarke-Groves)机制[16]是广义的单物品第二价拍卖。VCG 是一个 truthful 的机制，它的设计目标是最大化社会福利。

我们首先分析一般化的 VCG 拍卖。定义  $A = \{\text{分配方式集合}\}$ ，其中每个物品只能分配给一个买家，但一个买家可以获得多个物品。则 VCG 的分配规则为：选择一个分配方式  $a^* \in A$ ，使得社会福利  $\sum_{i=1}^n b_i(a^*)$  最大化。VCG 的支付规则为：买家  $i$  支付他对其他买家造成的负外部性的和：

$$p_i = \max_{a' \in A} \left( \sum_{j \neq i} b_j(a') \right) - \sum_{j \neq i} b_j(a^*)$$

**定理 2.6[6]**：在一个策略博弈中，如果  $\hat{a}_i$  是参与者  $i$  的弱优势策略，则我们将参与者  $i$  的收益变为

$$u'_i(a) = u_i(a) + h(a_{-i})$$

其中  $h(a_{-i})$  代表任意由其他参与者的策略决定的函数，则  $\hat{a}_i$  仍然是新博弈中的弱优势策略。

**证明：** 详见[6]

**定理 2.7[6]**：VCG 的分配与支付规则是 truthful 的。

**证明：** 对于买家  $i$ ，其效用函数为：

$$u_i = v_i(a^*) - \max_{a' \in A} \left( \sum_{j \neq i} b_j(a') \right) + \sum_{j \neq i} b_j(a^*)$$

其中  $a^*$  是 VCG 规则所给出的分配方式。根据前一个定理，我们可以忽略右式的中间项，因此有

$$u_i \sim v_i(a^*) + \sum_{j \neq i} b_j(a^*) = \text{社会福利}$$

根据  $a^*$  的定义，当其达到最大值时， $v_i = b_i$ ，因此虚假的报价无法使参与者  $i$  的收益提高。

在关键词拍卖中，VCG 与 GSP 机制的区别仅在于支付规则：每位买家需要支付的费用为他的出现给其他买家带来的负外部性。因此，最后一名得到广告位的

买家支付的费用与 GSP 相同：如果  $N \geq K$  则为 0，否则为  $\alpha_N b^{N+1}$ ；对于其他任意一个买家  $i$  ( $i < \min(N, K)$ )，支付的费用则为  $p^i = (\alpha_i - \alpha_{i+1})b^{i+1} + p^{i+1}$ 。

#### 2.4.4 阶梯式拍卖机制

阶梯式拍卖 (Laddered Auction) [4] 是一种 truthful 拍卖，它在保证 truthful 的同时仍能确保与 GSP 拍卖同等的盈利能力。

阶梯式拍卖的支付规则较为简单，不是一般性，假设买家  $i$  代表出价排在第  $i$  位的买家，对于任意的  $1 \leq i \leq \min(N, K)$ ，买家  $i$  所需支付的 PPC 为：

$$p^i = \sum_{j=i}^{\min(N, K)} \left( \frac{\alpha_{i,j} - \alpha_{i,j+1}}{\alpha_{i,i}} \right) \frac{w_{j+1}}{w_i} b_{j+1}$$

直观上来看：

1. 对于买家  $i$  如果在位置  $i+1$  上会收到的点击，他将会支付与  $i+1$  时同样的费用
2. 对于多出来的点击，买家  $i$  支付的费用为保持位置  $i$  所需要的最少费用

**定理 2.8[4]**：对于给定的权重  $w_1, \dots, w_n$ ，阶梯式拍卖是 truthful 拍卖。

**证明：**不失一般性，假设买家  $i$  即为排名第  $i$  的买假。考虑一个任意的买家  $M$ ，假设其他买家的出价一定，并且有  $p^{K+1} = 0$ 。注意到  $M$  最终支付的费用与其具体出价无关，而只与他的排名有关，因此我们只需考虑买家如何调整自己的排名。

设  $v_M$  表示买家  $M$  的真实股价，则  $M$  最终支付的价格为  $p^M$ ，那么显然有： $w_j b_j \geq$

$w_M v_M, \forall j < M$  以及  $w_j b_j \leq w_M v_M, \forall j > M$ 。要证明拍卖为 truthful 拍卖，只需证

明买家  $M$  不能通过谎报估价来提高自己的收益。我们采用反证法来证明。在所有能使买家  $M$  获利最大的排名中，设  $r$  为除  $M$  外离买家最近的广告位。现在假设买

家  $M$  能够通过谎报自己的估价来使自己排到  $r$ ，那么有两种情况：1.  $r > M$ ，此

时若买家选择排到  $r-1$ ，则能够进一步获利： $(\alpha_{M,r-1} - \alpha_{M,r}) \left( v_M - \frac{w_r}{w_M} b_r \right) \geq 0$ ；

2.  $r < M$ ，此时若买家选择排到  $r+1$ ，则能够进一步获利： $(\alpha_{M,r+1} - \alpha_{M,r}) \left( v_M - \right.$

$\frac{w_r}{w_M} b_r) \geq 0$ 。因此  $r$  不是获利最高的排名，矛盾。

**定理 2.9[4]**：阶梯式拍卖是加权降序排序拍卖中唯一的 truthful 拍卖。

**证明：**见[4]

## 2.5 保留价

在许多竞价拍卖中还存在保留价的设计，这一设计规定买家的出价不得低于某一个给定的值，否则竞价无效。保留价的设计可以在一定程度上保证广告位不会被买家以零或者接近于零的价格拍走，但另一方面却会潜在地降低买家参与的积极性，因为买家总会趋向于以保留价拍走物品。因此搜索引擎往往会将保留价对买家隐藏[20]。由于保留价的这种性质，如何设置保留价，以达到平台收益最大化就成为目前的研究方向之一。在 Hu[19]等的研究中，就讨论了第一价和第二价拍卖中的最优保留价的设置与风险规避问题。

## 3 第一价点击量拍卖 (FCA) 模型

在现实的搜索引擎中，每一个广告位获得的点击量都是一个可观的数字。然而实际的买家往往有自己期望的点击量，或者希望获得的点击量中的用户拥有某种特定的属性，这个点击量与搜索引擎提供的广告位的点击量可能相去甚远。为了满足广告主的需求，一种可行的办法是将广告位的概念弱化，而将点击量作为拍卖的主体，并按各个卖家的出价情况对点击量进行动态分配。在现实中，搜索引擎往往掌握着大量的广告资源，这样的动态调配方案是完全可以实现的，[7]中就曾提出过类似的概念。

一旦我们将不可分的广告位的拍卖转化为可分的点击量的拍卖，原有的拍卖模型便不再适用，我们需要寻找一个新的机制来满足我们的需求。事实上，前人已经提出了许多对连续可分的资源进行拍卖的机制，其中一种便是比例分配模型。这一模型在频谱分配[11]和计算资源拍卖[12]等问题上已经证明有着良好的表现。



据此，我们将点击量拍卖和按比例分配的优势结合起来，提出一种新的拍卖模型：第一价点击量拍卖(First-price Click Auction, FCA)模型。这个模型当中搜索引擎不再以广告位为单位，而是将已有广告位的点击量整合起来，统一进行拍卖，最后再通过动态的分配算法来将买家的广告放置到实际的广告位中，使其获得相应的期望点击量。

在继续讨论之前，先给出一些基本假设。

### 3.1 模型假设

1. 所有的买家都是理性并且自利的，目标都是追求自我收益最大化
2. 点通率都是可分(Separable)的
3. 不考虑买家的预算情况，即所有的买家的预算都是无限的
4. 不同用户在同一位置获得的点击量相同

其中 1, 2 都是常见的假设。3 和 4 是为了减小后文讨论难度所做的假设。事实上实际的拍卖中需要考虑的因素相当复杂，将每一个因素都考虑进模型几乎是不可能的，我们先通过简化假设的方法，得出简单情况下的结论，再在后续研究中进一步分析较复杂的情况。

值得注意的是，我们的假设中认为点通率是可分的，即存在  $\mu_1, \mu_2, \dots, \mu_n > 0$  和  $\theta_1 \geq \theta_2 \geq \dots \geq \theta_K > 0$ ，使得  $CTR_{i,j} = \mu_i \theta_j$ 。我们用  $\mathcal{L}_j$  表示第  $j$  的广告位在一段时间内的浏览量，则买家  $i$  的广告放置在第  $j$  个广告位上的点击量  $\alpha_{i,j}$  可以表示为：

$$\alpha_{i,j} = CTR_{i,j} \mathcal{L}_j = \mu_i \theta_j \mathcal{L}_j$$

其中  $\mu_i$  是由用户决定的量，而  $\theta_j$  和  $\mathcal{L}_j$  是由广告位决定的量。在本文中我们暂不讨论用户因素对于 CTR 的影响，不同用户在同一广告位时获得的预期点击量相同，即所有的  $\mu_i = 1$ 。因此在后文中，我们直接将  $\alpha_{i,j}$  写作  $\alpha_j$ 。

### 3.2 出价规则

出价规则与其他关键词拍卖机制相同，为连续拍卖，即买家可以在任意时间修改自己的出价，搜索引擎在获得新的出价以后动态调整点击量分配。

### 3.3 排序规则

我们在本节中给出两个用于将广告位转化为可重新分配的预期点击量的算法，为 3.5 中介绍的分配规则提供基础。这里假设搜索引擎已经根据所有买家的出价确定了广告主所获得的预期点击量  $c$ 。为了使买家获得期望的点击量，我们需要对广告主的广告进行动态的排序。所有买家最终拍得的点击量为  $c_1, \dots, c_{\min(N,K)}$ ，这些点击量必须满足：

$$\sum_{i=1}^{\min(N,K)} c_i = \sum_{j=1}^{\min(N,K)} \alpha_j$$

我们首先用算法 3.1 确定广告主的预期点击量在广告位上的分布：

---

**算法 3.1**：确定广告主所获的预期点击量在广告位上的分布

---

```

function solveRank(N,K,c,  $\alpha$ )
  m=1//m 代表当前正在分配的广告位
  n=1//n 代表当前正在分配的买家
  //g(i).rank(j)表示广告位 j 为买家 i 提供的预期点击量
  available_clicks = c(1)
  while m <= min(N,K) and n <= min(N,K)
    if available_clicks>=0// 使用 m 的点击量
      available_clicks -= c(n)
      if available_clicks>=0//m 的点击量未使用完
        g(n).rank(m) = c(n)
        n += 1
      else//m 的点击量已使用完
        g(n).rank(m) = available_clicks+c(n)
        m += 1
    endif
  else
    available_clicks += $\alpha$ (m)
    if(available_clicks>=0)//n 的点击量已满足

```

---

---

```

    g(n).rank(m) =  $\alpha(m) - \text{available\_clicks}$  // 广告位 m 提供 ( $\alpha(m) - \text{available\_clicks}$ ) 的点击量
    n += 1
else // n 的点击量未被满足
    g(n).rank(m) =  $\alpha(m)$  // 广告位 m 提供  $\alpha(m)$  的点击量
    m += 1
endif
endif
endwhile
return g

```

---

该算法的复杂度为  $O(\min(N, K))$ 。算法运行的结果是一组长度为  $\min(N, K)$  的 Hash 表数组  $g$ ，每个  $g(i)$  中的 hash 表项  $\text{rank}(j)$  保存了广告位  $j$  为买家  $i$  提供的预期点击量。显然，买家  $i$  获得的预期总点击量即为  $c(i) = \sum_j g(i).rank(j)$ 。需要注意的是，本算法是基于概率预期的分配算法，也就是说算法只指定了买家期望获得的点击量，而非准确的最终点击量。

**引理 3.1**：算法 3.1 为第  $i$  个买家分配的期望点击量为  $c_i$ 。

**证明**：首先证明  $\sum_{j=1}^{\min(N, K)} g(i).rank(j) = c(i), \forall i \in 1 \dots \min(N, K)$ 。由程序流程知  $\text{available\_clicks}$  首先减去  $c(n)$ ，随后每次为  $\text{available\_clicks}$  加  $\alpha(m)$  的同时也为  $g(n).rank$  加上  $\alpha(m)$ ，当  $\text{available\_clicks} \geq 0$  时，有  $\text{available\_clicks} = \alpha(1) + \dots + \alpha(m) - c(n)$ ，此时  $g(n)$  的所有  $\text{rank}$  的总和为  $\alpha(1) + \dots + \alpha(m+1)$ ，因此再为  $g(n)$  的  $\text{rank}$  添加  $\alpha(m) - \text{available\_clicks}$ ，则  $g(n).rank$  的总和变为： $\alpha(1) + \dots + \alpha(m+1) + \alpha(m) - \text{available\_clicks} = \alpha(1) + \dots + \alpha(m) - (\alpha(1) + \dots + \alpha(m) - c(n)) = c(n)$ 。

其次证明  $\sum_{i=1}^{\min(N, K)} g(i).rank(j) = \alpha(j), \forall j \in 1 \dots \min(N, K)$ 。注意在程序的控制流中，从  $m$  到  $m+1$  必然只经过一次  $\text{available\_clicks} += \alpha(m)$  和隐性的  $\text{if}(\text{available\_clicks} < 0)$ ，期间  $\text{available\_clicks}$  中所有来自  $\alpha(m)$  中的点击量全部都加到了某个  $g(i).rank(m)$  中，因此得证。

用算法 3.1 确定了数组  $g$  后，我们继续用算法 3.2 来为广告位选取广告。

**算法 3.2:** 根据广告主排名选取一次查询中的广告**function handleRequest(keyword)**

```

    g = getAdvertisers(keyword)//获取对 keyword 进行竞价的广告主数组 g
    for i = 1 to min(N,K)//遍历可分配的广告位
        candidates = []
        for j = 1 to min(N,K)//遍历有效的广告主
            if i in g(j).rank//如果广告位 i 为广告主 j 提供了点击量
                percentage = g(j).rank(i) /  $\alpha(i)$ //percentage 即为广告主 j 在广告位 i 中所占百分比
                candidates.append((j, percentage))//将(广告主,百分比)的二元组添加到候选列表中
            endif
        endfor
        select(i) = select_from_candidates(candidates)// 这里
        select_from_candidates 的具体操作省略, 其功能为: 以每个候选项占广告位 i 的百分比为概率, 选出并返回一个候选广告主的广告
    endfor
    return select

```

该算法的复杂度为 $O(\min(N, K)^2)$ 。算法返回一个长度为  $\min(N, K)$  的数组  $\text{select}$  作为结果。 $\text{select}(i)$  表示第  $i$  个广告位在本次拍卖中选择的广告主及相应的广告。

**定理 3.1 :** 算法 3.2 能够确保第  $i$  个买家最终期望得到的点击量为  $c_i$ 。

**证明:** 我们用随机变量  $X_{i,j}$  表示买家  $j$  在广告位  $i$  获得的点击量, 则买家  $j$  获得的总点击量为  $X_j = \sum_{i=1}^{\min(N,K)} X_{i,j}$ 。算法中买家  $j$  的广告在广告位  $i$  展示的概率为  $g(j).rank(i) / \alpha(i)$ , 因此  $E(X_{i,j}) = \frac{g(j).rank(i)}{\alpha(i)} * \alpha(i) = g(j).rank(i)$ , 由线性性可得  $E(X_j) = E\left(\sum_{i=1}^{\min(N,K)} X_{i,j}\right) = \sum_{i=1}^{\min(N,K)} E(X_{i,j}) = \sum_{i=1}^{\min(N,K)} g(j).rank(i)$ 。由引理 3.1 可得  $E(X_j) = c_j$ 。

### 3.4 支付规则

本拍卖采用第一价支付，即买家最终支付的价格为其出价。

### 3.5 分配规则

由于我们不再按照固定的点击量（即广告位）进行分配，如何让买家自行竞价，并获得最优的广告位就成为本机制设计的关键。这里我们提出分配函数的概念。分配函数的输入为用户的出价，输出则为其获得的点击量。我们将分配函数定义如下：

$$f: \mathbf{b} \rightarrow \mathbf{R}^K$$

$f_i(\mathbf{b})$ 表示出价 $b^i$ 的买家 $g^i$  ( $i \leq \min(N, K)$ ，即所有能获得广告位的买家)获得的点击量占总点击量的比例。显然， $f$  满足条件： $\sum_{i=1}^{\min(N, K)} f_i(\mathbf{b}) = 1$ ；

则我们可以将买家的效用函数定义为：

$$u_i(\mathbf{b}) = \mathcal{S} \cdot f_i(\mathbf{b}) \cdot (v_i - p_i) = \mathcal{S} \cdot f_i(\mathbf{b}) \cdot (v_i - b_i)$$

接下来主要关心不同  $f(i)$  下参与者的最适反应：

### 3.6 最适反应

#### 线性 FCA(LFCA)

我们以线性 FCA 为例分析买家的最适反应，分配函数为：

$$f_i(\mathbf{b}) = \frac{b_i}{\sum_{j=1}^{\min(N, K)} b_j}$$

即买家  $i$  分配到的点击量占总点击量的比等于其出价占总出价的比。

为分析第一个问题，我们将 $f_i(\mathbf{b})$ 带入 $u_i(\mathbf{b})$ ，再对 $u_i(\mathbf{b})$ 求 $b_i$ 的导，令求导后的结果等于 0，求极值点并得出最优出价 $br_i$ 。求得的结果为：

$$br_i = \sqrt{b_{-i}^2 + b_{-i}v_i - b_{-i}}$$

对这个结果进行分析可以得出两个结论，以下省略证明直接给出：第一， $v_i$

越高,  $br_i$ 越高; 第二,  $b_{-i}$ 越高,  $br_i$ 越高,  $b_{-i}$ 趋于无穷大时,  $br_i$ 趋于 $\frac{v_i}{2}$ 。也就是说, 无论如何买家最多只会出到估价的一半。

## 通用 FCA

通用分配函数定义为:

$$f_i(b) = \frac{h(b_i)}{\sum_{j=1}^{\min(N,K)} h(b_j)}$$

其中 $h(b_i)$ 为根据买家 $i$ 的出价 $b_i$ 换算得到的份额, 必须为严格单调递增函数。在线性模型中,  $h(b_i) = b_i$ 。同样我们也可以给出准线性模型 $h(b_i) = b_i + k$ , 多次模型 $h(b_i) = b_i^k$ , 指数模型 $h(b_i) = e^{b_i}$ 等等, 以调节模型的表现。

对通用 FCA 模型的最适反应分析与线性模型类似。如果我们要求出平衡时候的状态, 则需联立所有的最优解方程, 即解方程组:

$$\left\{ \begin{array}{l} \frac{\partial \left( s \cdot \frac{h(b_1)}{\sum_{j=1}^{\min(N,K)} h(b_j)} \cdot (v_1 - b_1) \right)}{\partial b_1} = 0 \\ \frac{\partial \left( s \cdot \frac{h(b_2)}{\sum_{j=1}^{\min(N,K)} h(b_j)} \cdot (v_2 - b_2) \right)}{\partial b_2} = 0 \\ \dots \dots \\ \dots \dots \\ \frac{\partial \left( s \cdot \frac{h(b_{\min(N,K)})}{\sum_{j=1}^{\min(N,K)} h(b_j)} \cdot (v_{\min(N,K)} - b_{\min(N,K)}) \right)}{\partial b_{\min(N,K)}} = 0 \end{array} \right.$$

这个方程组的求解显然是十分复杂的, 我们将在后文中用数值模拟的方法求出数值解, 并对均衡下的模型表现进行分析。同时, 我们将在后文中证明均衡的存在性, 从而上述方程可以证明是有解的。

### 3.7 纳什均衡存在性与唯一性

Rosen 的文章[14]中提出了凹博弈的概念：在博弈中，如果每个参与者的效用函数 $u_i(x)$ 在该参与者的策略空间 $x_i$ 内都是凹函数，则该博弈是凹博弈。如果所有参与者的效用函数都是严格的凹函数，则博弈是严格凹博弈。并且 Rosen 在该文中证明了如下定理，我们不加证明直接给出：

**定理 3.3[14]**：每一个  $n$  人凹博弈中都存在一个均衡点

**定理 3.4[14]**： $n$  人严格凹博弈中的均衡点唯一

根据以上定理我们可以证明 FCA 模型的均衡存在性。

**定理 3.5**：FCA 模型中存在均衡点，且均衡点唯一。

**证明**：首先对最简单的线性分配模型进行分析，对任意的买家  $i$ ，其效用函数为：

$$u_i(b) = \mathcal{S} \cdot \frac{b_i}{\sum_{j=1}^{\min(N,K)} b_j} \cdot (v_i - b_i)$$

$u_i$  的定义域为  $b_i > 0$ ，易证明在  $b_i > 0, \forall i = 1, \dots, \min(N, K)$ ， $\mathcal{S} > 0$  的条件下，有  $u_i'' < 0, \forall b_i > 0$ ，因此，在本机制中  $u_i$  在  $b_i > 0$  上为严格凹函数。由 Rosen 的定理 [14] 可得该博弈中存在均衡点，且均衡点唯一。

对于非线性的情况，即：

$$f_i(b) = \frac{h(b_i)}{\sum_{j=1}^{\min(N,K)} h(b_j)}$$

由于  $h$  为  $R^+$  上的严格单调递增函数，我们可以将买家的竞价做一个  $b_i$  到  $b'_i$  的一一映射：

$$b'_i = h(b_i)$$

则对于竞价为  $(b'_1, \dots, b'_{\min(N,K)})$  的博弈，由前面的证明可知存在唯一均衡，从而对于

于竞价为  $(b_1, \dots, b_{\min(N,K)})$  的博弈，也存在唯一均衡。

## 4 数值模拟

### 4.1 数据来源

我们的数值模拟采用 Yahoo! Labs [10]提供的来自真实交易中的关键词拍卖数据。其格式定义如下：

**Table 4.1** Yahoo! Labs Dataset Fields

**表 4.1** Yahoo! Labs 数据集表项

day	anonymized account_id	rank	anonymized keyphrase	avg bid	impressions	clicks
-----	-----------------------	------	----------------------	---------	-------------	--------

数据集中包括了 Yahoo!搜索平台中超过 4 个月的总共 77850272 条竞价条目。其中 day, account\_id, rank 和 keyphrase 为数据集主键，通其余项按照主键进行聚集。

### 4.2 数据预处理

在本文中，我们关注的是单次竞价中不同机制的表现，而不关心长期的变化，因此我们首先对数据集按照天数进行分类。其次，由于原始数据只给出了用户在一天内对某个关键词的出价信息，我们无法精确地还原某一次拍卖中的出价和支付情况，但我们可以将每个广告位的平均出价、点击量等信息提取出来作为该关键词的近似数据。因此，本文中采用的数据提取方法为：将单日的所有出价信息按照关键词->排名的方式进行分组，并对每一组中的所有出价信息求平均值，得到任意关键词在其所有广告位上的平均出价和平均点击量。通过这样的方式处理后，我们总共组合出 811765 次有效的关键词拍卖。

值得注意的是，由于 Yahoo! Labs 在提供这些数据时所采用的竞价机制为 GSP 机制，而 GSP 机制是非 truthful 机制，我们同样无法精确地得到每个用户的真实估价信息。但注意到，我们利用本数据集的目的并不是要重现现实的拍卖，而是获取到接近真实情况的竞价数据，从而对不同的竞价机制进行评估，因此我们不关心用户在当时的拍卖中是否 Truthful。为了简便起见，我们将用户的出价均视为 Truthful 出价。



### 4.3 模拟方法与评估标准

我们用 VCG 机制作为基线，用三次 FCA (Cubic FCA, CFCA) 模型，即分配函数

$f_i(b) = \frac{b_i^3}{\sum_{j=1}^{\min(N,K)} b_j^3}$ , 作为测试模型。对以下三个方面的指标进行评估：

1. 机制效率  $E$  (即社会福利)
2. 平台收益  $\mathcal{R}$  (即卖家收益)
3. 用户友好度  $\mathcal{F}$

其中每次拍卖中的机制效率定义为所有买家的广告所带来的收益之和，即：

$$E = \sum_{i=1}^{\min(N,K)} c(i) * v(i)$$

平台收益定义为所有买家支付的费用之和，即：

$$\mathcal{R} = \sum_{i=1}^{\min(N,K)} p(i)$$

用户友好度计算为广告主所获点击量与总支付的相关系数：

$$\mathcal{F} = \frac{cov(p, c)}{\sigma_p \sigma_c}$$

我们将前两项用 CFCA 模型所得值/基线模型所得值来表示，用户友好度用实际值表示。

### 4.4 总体分析

首先对总体情况进行统计，我们对机制效率和平台收益的分布进行统计，得到结果如下：

**Table 4.2** Comparison between CFCA and VCG Overview

**表 4.2** CFCA 与 VCG 对比总览

	机制效率总和( $\mathcal{E}$ )	平台收益总和( $\mathcal{R}$ )	相关系数平均值( $\mathcal{F}$ )
CFCA	4.9443e+10	3.3636e+10	0.9981
VCG	4.8976e+10	3.5299e+10	0.9966

注意到 FCA 模型在机制效率总和 $\mathcal{E}$ 相关系数 $\mathcal{F}$ 上均优于基线机制，前者说明 FCA 模型在最大化总体资源分配效率上表现更优，后者表明 FCA 模型对个体的资源分配方式略微占优，因为买家获得的点击量份额与其出价更为相关。但与此同时 FCA 模型的平台收益总和却略低于 VCG。综合来看，FCA 模型创造了更多的社会福利，但却收取更少的利润，同时使得用户的投资回报更优。因此我们可以认为 FCA 模型是一个对用户更加友好的模型。

我们对每次交易中，FCA 模型的各项数据对基线模型 VCG 的比值进行分析。值得注意的是，这里进行统计的是以单次交易为单位的总体对比情况，而非对全部交易的总体情况进行对比。因此在这里我们关注的是交易的分布情况。

结果如下：

效率比值分布：

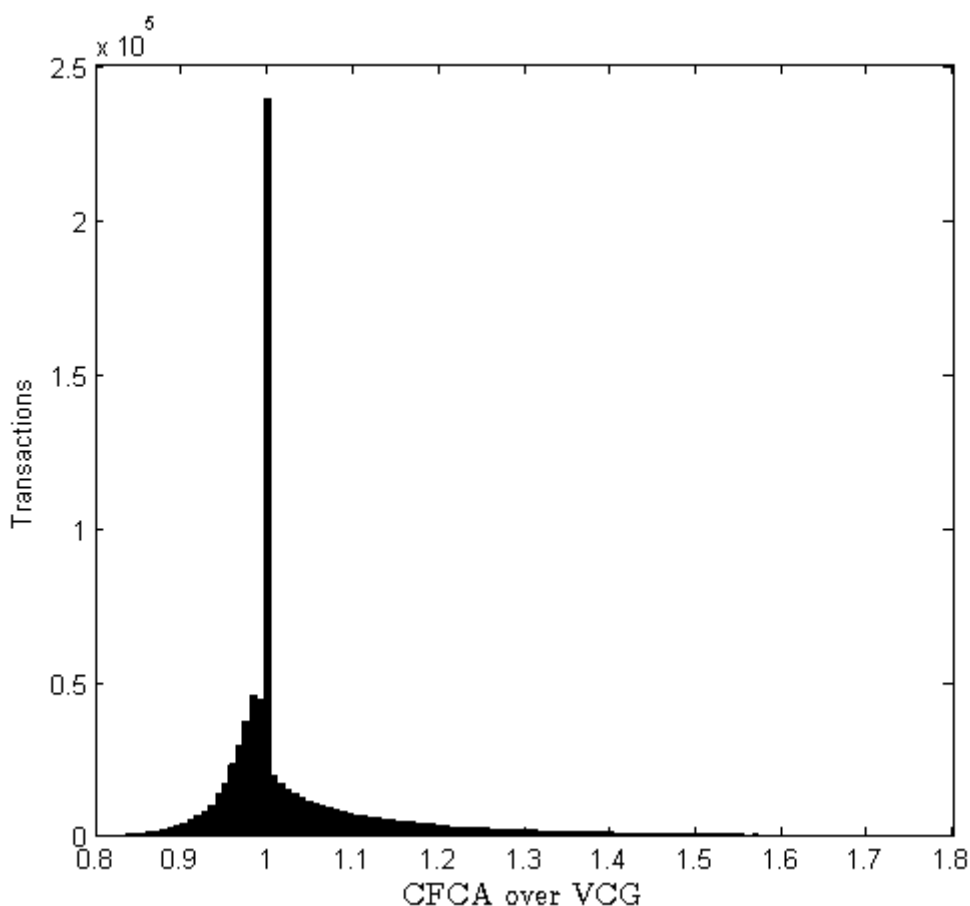


Fig. 4.1 Distribution of CFCA's Efficiency over VCG

图 4.1 CFCA 与 VCG 效率比值分布

可以看出 FCA 模型在机制效率上基本与 VCG 机制持平，绝大多数的交易的效

率与基线机制相当。进一步从数值上进行统计，可得：

Table 4.3 Overview of CFCA’s Efficiency over VCG

表 4.3 CFCA 模型与 VCG 效率比值总览

	均值	标准差	中位数	众数
效率比值	1.0648	0.1986	1.0000	1.0000

注意到中位数和众数均为 1.0000，而比值的变化系数(coefficient of variation)为 $CV = \frac{\text{标准差}}{\text{均值}} = 0.1865$ 。因此可以看出，FCA 模型相比 VCG 在效率上几乎完全相同，而在均值上微弱领先于 VCG。

平台收益比值分布：

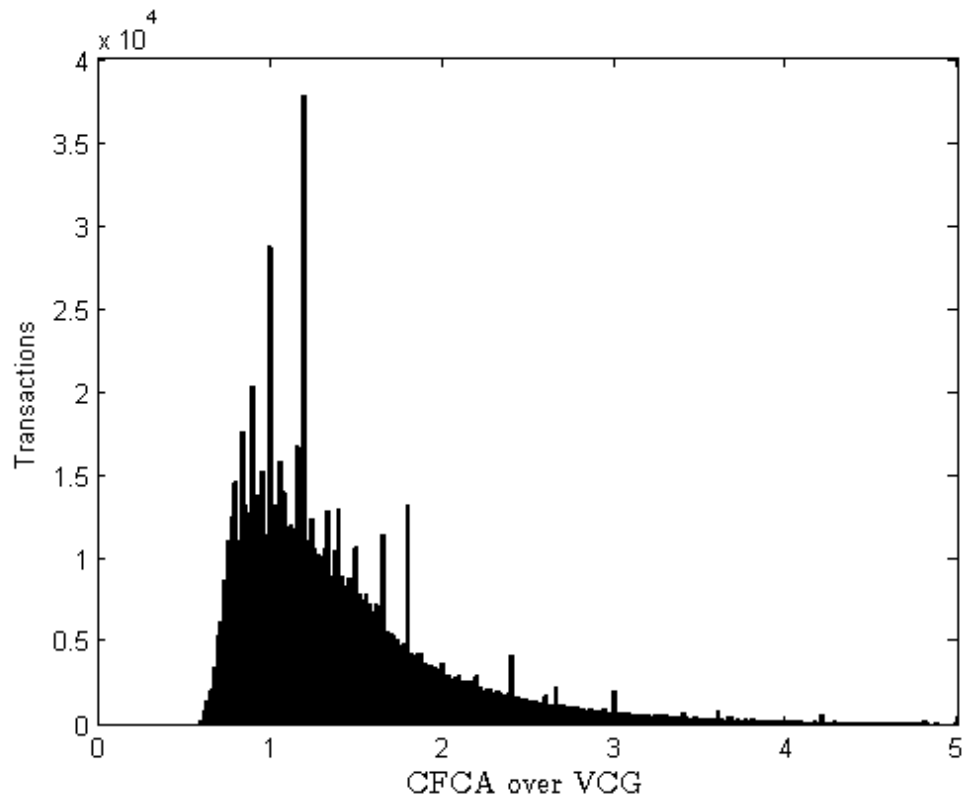


Fig. 4.2 Distribution of CFCA’s Platform Revenue over VCG

图 4.2 CFCA 与 VCG 平台收益比值分布

可以看出 FCA 模型在少部分情况下收益低于 VCG，大部分收益在 VCG 的 1-3 倍。

从数值上进行统计则有：

Table 4.4 Distribution of CFCA’s Platform Revenue over VCG

表 4.4 CFCA 模型与 VCG 平台收益比值分布

	< 1.0	1.0 ~ 2.0	2.0 ~ 3.0	> 3.0
平台收益比值	24.66%	58.85%	11.09%	5.40%

由此可以看出 FCA 模型在多数交易中的收益大于 VCG，但由前面的分析可知其总收益略低于 VCG 机制。这是因为 FCA 模型在少数交易 (24.66%) 中表现不佳所致，我们将在后文中进一步分析这些交易的特点。

相关系数分布对比：

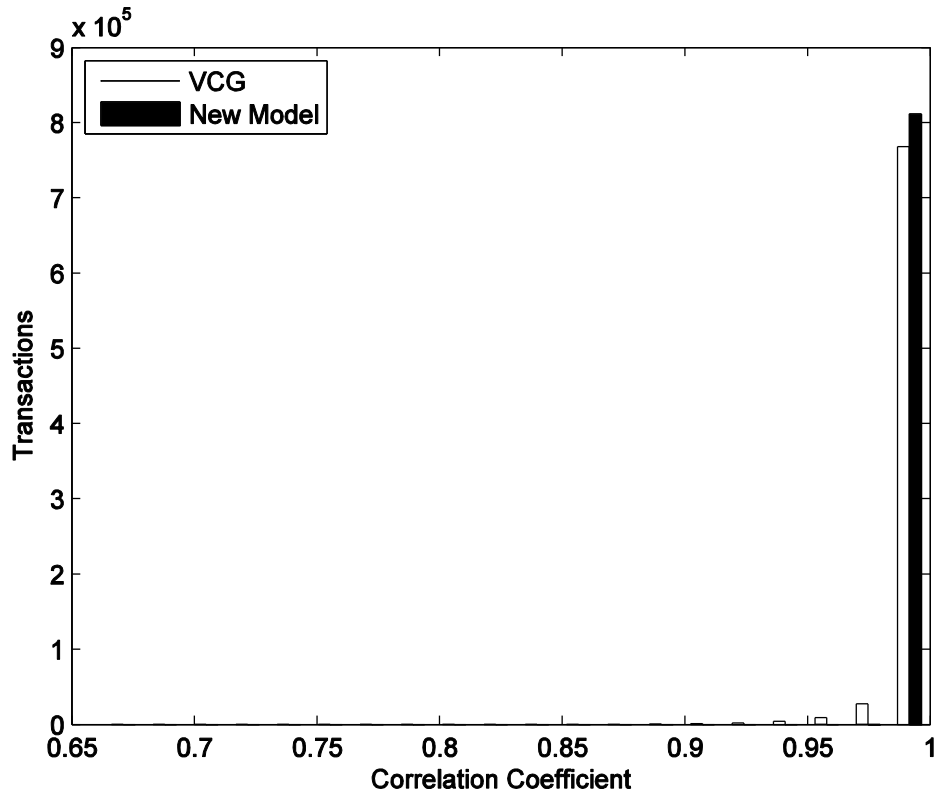


Fig. 4.3 Distribution of Correlation Coefficient of Advertisers' Share and Payment

图 4.3 广告主点击量份额与支付相关系数分布

可以看出 FCA 模型的相关度完全集中于接近 1 的区间，而 VCG 仍存在相关度较低的部分。事实上，FCA 模型的最低相关度为：0.9759，均值为 0.9981，好于 VCG 的 0.6627 和 0.9966。从用户的角度来看，绝大多数用户在 FCA 模型中都获得了与投资相匹配的回报，而 VCG 模型在少数交易中用户的支付与回报相差较大。

## 4.5 有效竞价人数对 FCA 模型的影响

我们以有效广告位（等同于有效竞价人数）为横坐标，FCA 模型与 VCG 的值为纵坐标作出密度图：

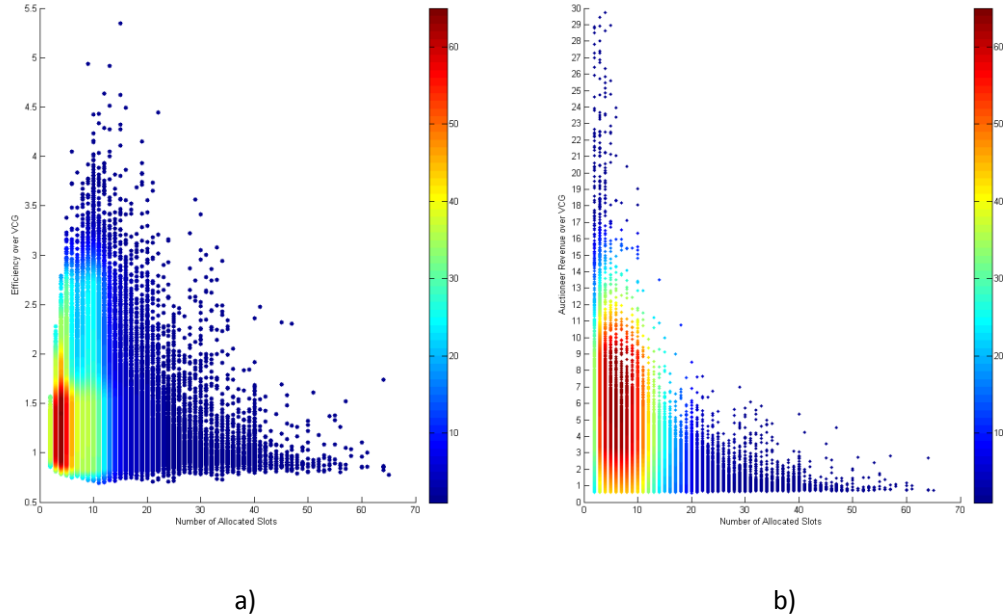


Fig. 4.4 a) Impact of Number of Slots on CFCA's Efficiency over VCG  
b) Impact of Number of Slots on CFCA's Revenue over VCG

图 4.4 a) 广告位个数对模型效率比分布的影响  
b) 广告位个数对平台收益比分布的影响

可以看出 FCA 模型在有效广告位较少时（0-10 个）效率比分布逐渐扩大，在超过 10 个广告位后逐渐收窄，并趋于与 VCG 的效率相同。平台收益方面，随着有效广告位增多，FCA 模型的效率逐渐趋于与 VCG 机制相同。并且可以看出 FCA 模型在有效广告位较少时，平台收益达到 VCG 平台的 1-10 倍，收益能力十分可观。

值得注意的是，FCA 模型的高效率和低效率的范围几乎是同时扩大和收窄的，但总体而言高效率的分布范围比低效率更广。

总体而言，我们可以得出的结论是 FCA 模型在有效广告位较少时表现超过 VCG，而在有效广告位较多时接近 VCG。

## 4.6 点击量分布对 FCA 模型的影响

我们以单次拍卖中所有广告位点击量的标准差为横坐标，FCA 模型与 VCG 的值之比为纵坐标作出密度图：

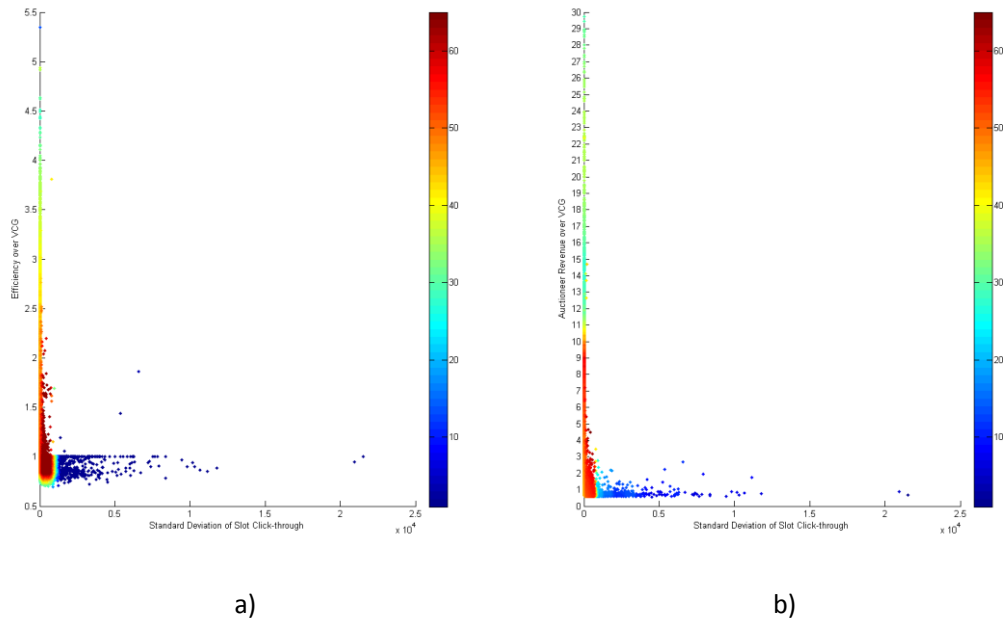


Fig. 4.5 a) Impact of Standard Deviation of Click-through's on CFCA's Efficiency over VCG  
b) Impact of Standard Deviation of Click-through's on CFCA's Revenue over VCG

图 4.5 a) 点击量标准差对模型效率比分布的影响  
b) 点击量标准差对平台收益比分布的影响

FCA 模型中高效率的交易几乎全部分布在点击量标准差接近于 0 的区域，而随着标准差的增大，FCA 模型的机制效率变为与 VCG 机制相同或低于 VCG 机制。平台收益方面，高收益的交易同样集中在点击量标准差接近于 0 的区域，随着标准差的增大，平台收益同样趋近于 VCG 机制。

因此从测试结果来看，FCA 模型在广告位较少，广告位点击量标准差较小的情况下，多数交易表现好于 VCG，而在广告位增多，点击量标准差增大时，表现趋近于 VCG。

## 4.7 模型迭代次数

FCA 模型是一个重复博弈模型。随着博弈的进行，参与者获取的信息越来越多，他们将会并不断改变自身策略。如果模型存在均衡，则最终经过一定回合的博弈之后，所有买家的策略将会收敛到一个平衡点。我们在关注模型最终表现的同时，还需要关注模型达到平衡的速度，即模型达到平衡所需的迭代次数。如果迭代次数较高，在实际应用中就可能难以达到平衡。我们对所有交易中的迭代次数进行统计，得到：

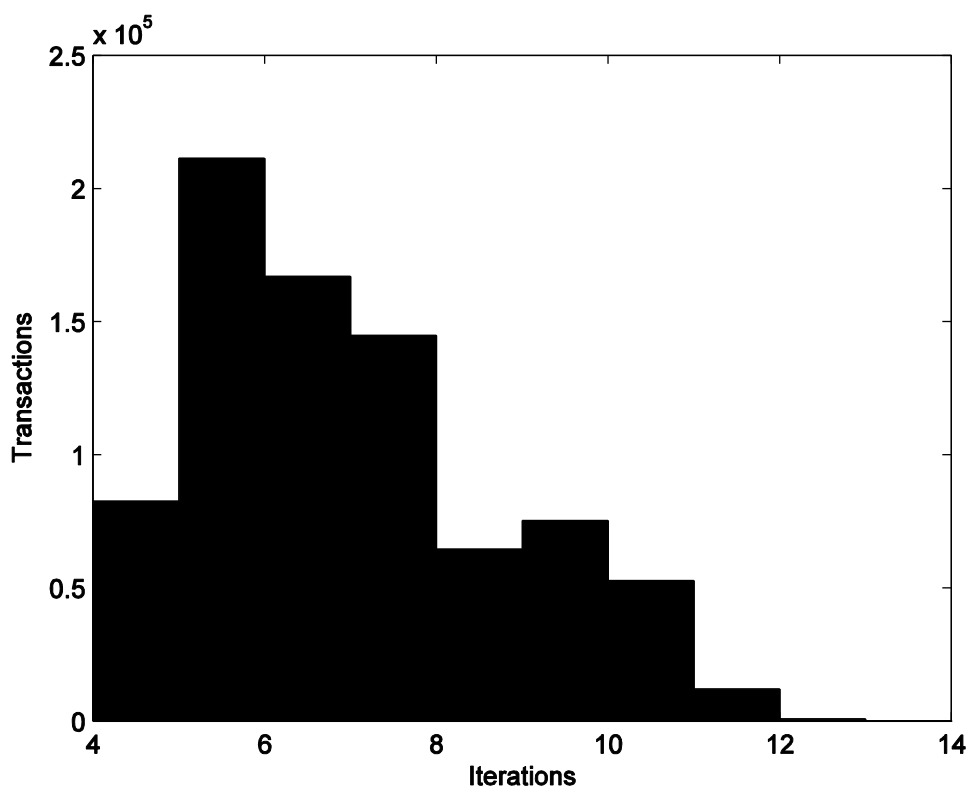


Fig. 4.6 Distribution of CFCA's Iterations before Reaching Equilibrium

图 4.6 CFCA 达到均衡前的迭代次数分布

从统计图中可以看出，所有的交易均在 14 次以内就达到了平衡，最少经过 4 次迭代达到平衡。考虑到现实的关键词拍卖中出价十分频繁，因此我们认为 FCA 模型能够在较短的时间内达到平衡。

## 4.8 广告位重叠

由于模型对点击量采取按比例分配的方式，买家可能会获得来自多个广告位

的点击量。我们对每次交易中，最多覆盖的广告位数进行分析，得到：

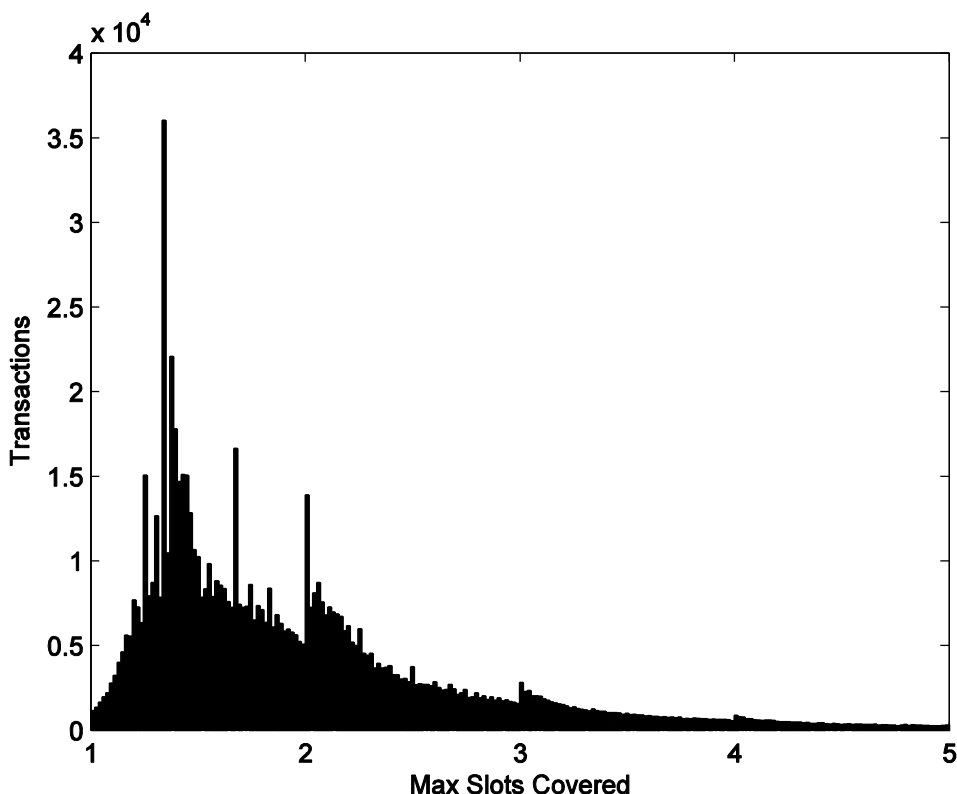


Fig. 4.7 Distribution of Max Covered Slots in CFCA's Transactions over VCG

图 4.7 CFCA 交易中最大覆盖广告位分布

大多数交易中单个买家最多覆盖的广告位在 3 个以内。事实上，当覆盖的广告位超过 1 时，就会出现某次检索中两个位置显示同一个广告的情况。这种情况下广告的点通过率会如何变化，从现有的研究来看尚不可知，可以作为未来研究的方向。如果造成了点通过率下降，那么我们需要对 FCA 模型的分配机制做进一步的优化。

## 4.9 极端情况分析

分析我们的测试结果可以发现，在部分拍卖中 FCA 模型的平台收益要远远大于 VCG 的平台收益，其中差距最大的甚至达到了 276 倍，我们对该极端个例进行分析。

在该次拍卖中，广告位的点击量和有效竞拍人的估价分别为：

Click\_through= [878, 872]



Valuation = [370, 150]

我们将机制效率、平台收益、以及两名买家的支付与收益等值列举如下：

**Table 4.5** Transaction Detail in the Extreme Case

**表 4.5** 极端案例中的交易细节

	机制效率	平台收益	买家 1 支付	买家 1 收益	买家 2 支付	买家 2 收益
FCA	5.5605e+05	2.4881e+05	2.0543e+05	2.8826e+05	4.3387e+04	1.8968e+04
VCG	4.5566e+05	900	900	3.2396e+05	0	1.3080e+05

在本次拍卖中 VCG 的平台收益只有 900，而 FCA 模型的平台收益达到了  $2.4881e+05$ 。分析买家的支付可以发现，根据 VCG 的支付规则，买家 1 只需支付  $(878-872)*150=900$ ，而买家 2 则不需支付，这也导致两位买家的支付与收益相差悬殊。而在 FCA 模型中，虽然两名买家的收益变少，但与其支付的价格的比例更为合理。

在本次拍卖中，造成 FCA 模型与 VCG 收益相差巨大有多方面的原因：

首先，VCG 机制不适合在点击量不可分 (Separable) 时使用，这一点 Aggarwal 等 [4] 已经在其论文中给出了相应阐述。在我们的实验数据中，绝大部分点击量都是可分的，因此可以将这类情况视为少数个例。

其次，现实中的搜索引擎广告位的点击量是呈近似指数分布的。Feng [3] 等对 Yahoo、MSN、AltaVista 等搜索引擎的点击量数据进行统计得出，广告位从高到低的点击量分布相当好地符合系数为  $\delta = 1.428$  的指数衰减，即广告  $j$  放置在位置  $i$  时的点击率为  $\frac{\alpha_j}{\delta^{i-1}}$ 。而在本实验的设置中，由于点击量数据通过同一位置在一天中的平均值计算得到，因此在竞价人数较少时容易出现不同位置点击量雷同的情况。

最后，现实中的搜索引擎通常都采用保留价机制，对于每个位置有最少出价的要求，这样就避免了某些位置用户收益与支付极为不平衡的情况。

## 5 结论

本文中我们在点击量拍卖和比例分配模型的基础上，提出了一种新的关键词拍卖模型：第一价点击量拍卖模型。我们首先给出了用于动态分配点击量的基础算法，从而将不可分的广告位转化为可分的期望点击量。在此基础上，我们提出

了一套基于比例分配模型的竞价机制,从理论上分析了模型中买家的最适反应等竞价行为,并证明了模型存在唯一的纳什均衡。进一步,我们利用 Yahoo! Labs 提供的竞价数据,以 VCG 机制为基线机制,对 FCA 模型进行了评估。并从模型总体表现和不同因素的影响等方面进行了分析。

通过理论分析与实验验证,我们得出 FCA 模型在机制效率和用户友好度上均优于 VCG 机制,但在平台总收益上略低于 VCG 机制。总体来看,FCA 模型通过对广告位进行动态分配,并对预期点击量进行比例分配,实现了预期点击量拍卖,并在减少极少的平台收益的基础上,提高了平台的总体效率和用户友好度。

此外,本模型还存在许多局限性。首先,模型假设中没有设置买家的预算。但在现实中,买家的广告预算是有上限的。将买家预算考虑进模型后,模型中买家的策略以及最终的结果往往不同。Feldman[18],Hafalir[21]就对有预算限制下的买家最优化竞价策略进行了探讨,并对 GSP 和 VCG 下的情况分别进行了分析。在 Feldman 的最优策略中,如果买家的预算超支,则将会降低出价。而在本文探讨的模型中,没有对买家的预算作出限制,这是今后继续探讨的方向之一。

其次,模型在处理竞价数据时,由于数据集本身的局限性,只能对一天当中的竞价数据进行平均后重组近似的拍卖数据。然而现实当中许多关键词的竞价频率很高,某次竞价当中的参与者与全天的参与者不尽相同,报价数据自然也不同。因此本文中的实验得到的是近似结果而非精确结果。如果要进一步分析模型在真实情况下的表现,我们需要更加精确与详尽的数据集进行测试。

第三,FCA 模型假设任意买家在同一位置所获得的预期点击量相同,忽略了用户因素对于点击量的影响。在现实的搜索引擎中,不同的广告放置在同一位置获得的点通率不尽相同。如果将用户因素 $\mu_i$ 考虑进 FCA 模型,则分配算法和分配机制都会有所改变,这也是今后研究的方向之一。

最后,尽管 FCA 模型将固定的广告位转化为可分的点击量,在实际的分配中仍然会遇到“广告位重叠”的现象,即某一买家的广告可能在两个或多个广告位同时出现。这样的分配方式会对实际的点通率(click-through-rate)产生正面还是负面的影响,从现有的研究来看尚不可知。因此,在未来的研究中,我们需要对广告位重叠时点通率的变化做进一步的分析。如果点通率上升,我们可以适当改进分配算法,使得重叠率上升;如果下降,那么我们需要进一步研究如何在 FCA

模型中避免广告位重叠的现象发生

## 参考文献

- [1] Edelman, Benjamin, Michael Ostrovsky, and Michael Schwarz. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. No. w11765. National Bureau of Economic Research, (2005).
- [2] Jansen, Bernard J., and Tracy Mullen. "Sponsored search: an overview of the concept, history, and technology." *International Journal of Electronic Business* 6.2 (2008): 114-131.
- [3] Feng, Juan, Hemant K. Bhargava, and David M. Pennock. "Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms." *INFORMS Journal on Computing* 19.1 (2007): 137-148.
- [4] Aggarwal, Gagan, Ashish Goel, and Rajeev Motwani. "Truthful auctions for pricing search keywords." *Proceedings of the 7th ACM conference on Electronic commerce*. ACM, (2006).
- [5] Interactive Advertising Bureau/PriceWaterhouseCoopers, "IAB Internet Advertising Report: 2013 Full Year Results," (April 2014). [www.iab.net](http://www.iab.net).
- [6] Pritchard. Lecture. Game Theory and Algorithms. EPFL, Lausanne. 5 April. (2011). Lecture.
- [7] Kristanto, Ferry. "An allocation approach of sponsored search auctions." (2012).
- [8] "Lambert W function." Wikipedia, The Free Encyclopedia. Wikimedia Foundation, Inc. (February 2014). Web. (February 2014).
- [9] Borgers, Tilman, et al. "Equilibrium bids in sponsored search

---

auctions: Theory and evidence." (2007).

[10] Yahoo! Webscope. Yahoo! Search Marketing Advertiser Bid-Impression-Click data on competing Keywords. version 1.0. Yahoo! Labs.

[11] Huang, Jianwei, Randall A. Berry, and Michael L. Honig. "Auction-based spectrum sharing." *Mobile Networks and Applications* 11.3 (2006): 405-418.

[12] Feldman, Michal, Kevin Lai, and Li Zhang. "The proportional-share allocation market for computational resources." *Parallel and Distributed Systems, IEEE Transactions on* 20.8 (2009): 1075-1088.

[13] Aggarwal, Gagan, Jon Feldman, and S. Muthukrishnan. "Bidding to the top: VCG and equilibria of position-based auctions." *Approximation and Online Algorithms*. Springer Berlin Heidelberg, (2007). 15-28.

[14] Rosen, J. Ben. "Existence and uniqueness of equilibrium points for concave n-person games." *Econometrica: Journal of the Econometric Society* (1965): 520-534.

[15] Nash, John F. "Equilibrium points in n-person games." *Proceedings of the national academy of sciences* 36.1 (1950): 48-49.

[16] Vickrey, William. "Counterspeculation, auctions, and competitive sealed tenders." *The Journal of finance* 16.1 (1961): 8-37.

[17] Fudenberg, D. ; Tirole, Jean (1991), *Game Theory*, Cambridge MA: MIT Press.

[18] Feldman, Jon, et al. "Budget optimization in search-based advertising auctions." *Proceedings of the 8th ACM conference on Electronic commerce*. ACM, (2007).

[19] Hu, Audrey, Steven A. Matthews, and Liang Zou. "Risk aversion and optimal reserve prices in first-and second-price auctions." *Journal of Economic Theory* 145.3 (2010): 1188-1202.

[20] Vincent, Daniel R. "Bidding off the wall: Why reserve prices may be kept secret." *Journal of Economic Theory* 65.2 (1995): 575-584.

---

[21] Hafalir, Isa E., R. Ravi, and Amin Sayedi. "A near Pareto optimal auction with budget constraints." *Games and Economic Behavior* 74.2 (2012): 699–708.

## 致谢

首先感谢我的指导老师吴骏老师，感谢他在繁忙的科研工作中抽空来帮助我分析模型，对我论文的发展进行引导，并在一些重要的问题上给予我指导。同时也感谢俊元老师、王崇骏老师对本论文进展的严格把关、悉心关注，使我始终保持研究的热情。在老师们的指导与帮助下，我系统地学习了博弈论以及付费搜索这两个陌生领域的知识，阅读了大量文献，并对大量的数据进行了分析与处理。这个过程中我收获的不只是专业知识，还有沟通交流的技巧、阅读提炼文献的能力以及严谨客观的科研精神，这些都是我人生中的宝贵财富。

感谢我的同学和朋友们对我的鼓励与建议。论文写作是一次艰苦的历程，不仅要反复研读大量的文献，还要进行严密的理论推导和实验论证。正是因为有你们的陪伴，我才有动力战胜重重困难，最终顺利完成论文。感谢父母对我的关心与照顾，使我始终能够保持良好的状态来从事论文的写作。

最后，感谢百忙之中抽出时间对本文进行评审的老师。我深知自身学问尚浅，希望老师们对论文的不足之处不吝指正，你们的意见将是我未来不断努力的方向！