

文章编号:1000-5641(2013)03-0060-10

考虑信誉的广告投放机制设计及投放效率测度

殷红¹, 许彦如², 王长波²

(1. 华东师范大学商学院, 上海 200062; 2. 华东师范大学软件学院, 上海 200062)

摘要: 网络广告作为一种新兴广告形式正在高速发展. 如何得到最大化的广告收益是最受关注的问题. 但是, 单纯追求点击价值并不一定能给广告商带来最大的收益, 还可能带来网络欺诈、虚假广告等道德风险问题. 因而, 广告主的信誉越来越成为广告被关注和点击的重要因素. 本文在考虑广告主信誉、广告商品属性匹配度以及点击价值等因素的基础上, 基于机制设计理论, 给出了一种能实现广告整体收益最优的投放策略决策方法, 并进一步探讨了广告投放的效率测度. 最后, 通过可视化分析以及仿真实验来说明本文决策方法所给出的机制在投放效率方面的优越性.

关键词: 网络广告; 机制设计; 广告主信誉; 投放效率测度

中图分类号: TP391, C391 **文献标识码:** A **DOI:**10.3969/j.issn.1000-5641.2013.03.007

Research on the mechanism design and measure of quality for online advertising in consideration of reputation

YIN Hong¹, XU Yan-ru², WANG Chang-bo²

(1. Business School, East China Normal University, Shanghai 200062, China;

2. Software Engineering Institute, East China Normal University, Shanghai 200062, China)

Abstract: How to get the maximum profit with online advertising is the most concern issue for advertisers. However, only pursuing the Click Through Rate maybe can not bring the maximum profit, which even induces huge amounts of network fraud information and fraudulent transaction. Considering the factors of the advertisers' reputation, matching degree of product and clicking value, based on the mechanism design theory, an optimal launching mechanism for online advertising was proposed here. Further, the quality measure of advertise launching was calculated based on our mechanism. Through the final simulation testing and visual analysis, it was shown that our mechanism considering reputation has superiority on the launching efficiency.

Key words: online advertising; mechanism design; advertiser reputation; measure of quality

收稿日期:2013-03

基金项目:上海市教委科研创新项目(12ZZ042);国家社会科学基金(10CGL014)

第一作者:殷红,女,博士,副教授,研究方向为博弈论、系统工程等. E-mail: hyin@jjx.ecnu.edu.cn.

通信作者:王长波,男,教授,研究方向为计算机仿真、数据可视分析. E-mail: cbwangcg@gmail.com.

0 引言

伴随着互联网经济的快速发展,近年来更加符合用户体验模式的网络广告得到了迅猛的发展,市场占有率逐年提高.然而,随着网络广告越来越多,用户逐渐产生了“免疫”,致使近年来点击率(Click Through Rate, CRT)有所下降.因此,如何提高广告的投放效益问题成为近年的研究热点.

广告的投放效率与多个因素相关,其中定向的准确性是最为关键的因素.大部分网络广告系统使用基于用户的定向:用户被按照地理位置、浏览器设置、导航习惯、注册的信息进行识别和分类;然后进行广告的个性化定向,包括统计、优化和调度模型^[1].但是,单纯追求点击率及短期收益并不一定能给广告商带来最大的收益.相对于传统的广告,网络广告存在更大程度的信息不对称,因而更易诱发各类道德风险和网络欺诈,其中商家为增进自身利益而侵犯消费者合法权益的道德风险问题最为突出,如虚假广告、伪劣产品、信用炒作,以及近年来屡屡爆出的搜索引擎广告黑幕等.长远来看,信誉也是影响广告整体投放效益的重要因素.

网络广告的泛滥使接收能力有限的消费者的注意力成为“稀缺资源”,如何最有效率地分配这些稀缺资源成为所有广告主和广告公司最关心的重要问题.因此,设计一套最优的投放机制使各方的利益最大化,实现资源的有效配置,并研究此机制下的广告投放效益预测具有重要的研究意义.

本文在考虑广告主信誉、广告商品属性与用户偏好匹配度以及点击价值等因素的基础上,基于机制设计理论研究了广告投放的最优机制问题.具体的研究思路如下:首先探讨基于用户偏好的广告匹配度计算以及广告信誉的综合评分计算;然后在机制设计的框架下给出广告投放机制设计模型,通过模型求解得到最优的广告投放机制;最后通过对广告投放效率的测度定义以及仿真实验,来说明本文所设计机制在投放效率方面的优越性.

1 相关工作

相关的工作主要集中于网络广告的用户定向、信誉计算、机制设计及测度定义等方面.

对于网络用户购物的行为和偏好一直是数据挖掘工作的热点.近年来,研究工作从已有的内容定向、频率定向、时间定向、地域定向,扩展至以用户特征为基础的定向功能.罗枫^[2]认为可对网络媒介(网站规模、内容、性质等)进行细分,抓住消费者的需求,实行个性化的广告投放,实现资源利用的最优化和取得利益的最大化.孙涛^[3]进一步提出了一种综合 Web 页面内容、历史会话记录,以及对用户当前会话中的行为进行跟踪和挖掘的混合个性化定向模式^[3].针对用户的行为定向,如何结合商品的属性关联规则进行匹配是个难题.很多学者使用协同过滤模型^[4]或关联规则^[5]等推荐算法;Lee^[6]和 Papagelis^[7]则等使用聚类技术大大缩小协同过滤中用户和最近邻居搜索范围,提高搜索效率;Han^[8]等人提出新的关联计算方法,避免 Agrawal 提出算法中频繁集计算的缺陷.

目前对于网络广告的信誉研究还不多.已有的研究工作主要集中在商家的信用评价体系上,建立了信誉模型、累加模型、均值模型等不同的衡量体系.如朱艳春^[9]等找出了影响网上交易信任的五项因素,分别为交易价值、评分用户信任度、时间权重、信誉反馈评分和近期信誉信任度;李维安等^[10]针对淘宝数据和信誉评价系统研究了网上交易的声誉机制;张先

锋等^[11]分析了卖家在长短不同时期内肯定信用度和否定信用度对其销售价格、售出概率、售出数量,以及售出价格的影响,同时引入了运费、交易剩余时间和消费者保障计划等控制因素。尽管信誉评价体系越来越丰富,但却引入了更多的评判参数,难以被方便快捷地使用。

在对广告的投放机制研究方面,除了针对用户定向的个性化投放策略研究外,还有一部分研究集中于研究搜索广告位的拍卖机制问题。目前业界普遍采用的是广义第二价格(GSP)拍卖机制。Edelman等^[12]以Yahoo!为研究对象建立了付费搜索关键字广告的GSP拍卖模型,定义了一类局部无嫉妒均衡。同时,Varian^[13]以Google的拍卖规则为例建立了位置拍卖模型,并给出了对称Nash均衡的实现条件。基于机制设计理论来研究广告拍卖机制设计问题的文献并不多见:Garg等^[14]在这方面做了一些工作,刘刚等^[19]提出了一种基于星形用户社区模型的Twitter广告最优投放机制;然而他们并未考虑广告主的信誉因素。对于互联网定向广告,影响广告效益的因素不同,采用的机制也需要进一步研究。

研究广告投放策略的目的是希望达到最好的广告投放效率。当前对于广告投放效率计算主要采用点击率预测,如卫强对网络展示广告位置点击率进行的实证研究^[15],童强对搜索引擎关键字的点击率进行的预测研究^[16]。广告投放效率是就用户搜索目标而言所投放广告为满足其偏好所提供的便利程度,信息距离理论中的信息代价可以用来衡量用户为辨识目标信息所花费的资源。因此基于信息代价来对广告投放效率进行测度会更加合理,另外广告收益也是衡量广告投放效果的一个很重要的维度。本文也研究了基于广告收益的测度方法。

2 基于用户偏好的广告匹配度计算

广告投放过程中,用户个性偏好对广告投放效果有很大影响,因此必须计算待投放广告商品与用户偏好的匹配程度。考虑到商品的属性可能很多,属性的重要程度也不一样,这里提出一种基于实体抽象的用户偏好匹配度计算模型。具体来说,即借鉴数据库中实体抽象的思想,在复杂的商品属性和难以确定的用户之间建立层次关系,并根据该层次关系找出符合用户偏好的潜在广告商品。

以手机广告为例,假如需要对某一正在浏览手机相关信息的用户投放手机广告。用户可能感兴趣的手机类型跟其偏好有关;而不同型号的手机在价格、外观、配置等属性上具有迥异的表现。对于一般用户来说,多数属性并非其所熟悉的,且种类属性数量众多,关心每一个属性是否满足其偏好是不现实的。根据手机属性总结和用户选择的聚类,发现一个有趣的现象:众多的属性参数不过是一个具体信息,而用户的选择聚类则是一个抽象的过程。因此,借鉴关系数据库中的实体关系理论来简化对用户需求的判断,将商品的所有属性看做一个商品实体的汇聚,而将抽象的用户选择看做用户实体的属性。

另一方面,根据一定数量的用户使用手机的统计聚类发现,用户对手机的需求有很大差别。手机用户的选择汇聚成了几个类别,把这几个类别和数据对比可以将其分为几个用户的需求类型,例如商务类型、手机发烧类型等;而这些类型则可以抽象成用户的实体属性。得出的手机实体和用户实体存在对应关系,也就是用户实体与手机商品广告之间的对应关系,如图1所示。

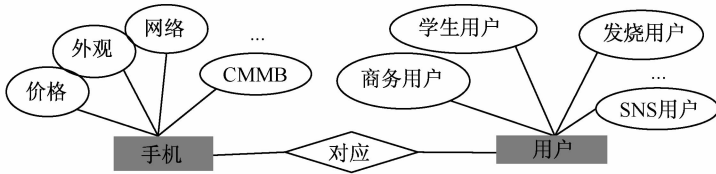


图1 手机商品实体与用户实体关系

Fig. 1 The relationship between the entity of mobile phones and users

这样,通过聚类分析可总结出手机用户类型及其满足的参数值大小,要计算某型号手机广告与手机用户类型的匹配程度,只需计算参数条件满足数占所有参数的比例值即可。

3 基于因子分析的广告信誉综合评分

广告商在决定所投放的广告类型时,会考虑广告商品与网页浏览用户偏好的匹配度,以及广告主对用户每点击价值(Click-Through Value)的报价。然而,在匹配度和报价相近的情况下,口碑越好的广告商品越容易引起用户的关注和点击,因此广告主的信誉也是影响用户点击率进而影响广告投放效率的一个很重要的因素。

对于广告主的信誉值,一些交易网站都提供关于其商品交易情况的信誉评价。比如淘宝网中商家的信誉评价体系分为卖家信息、店铺服务、店铺半年内动态评分、店铺30天内服务情况、卖家信用评价展示和卖家历史信用等几个方面。信誉评价体系虽然能从各个方面对商家的交易诚信和售后服务诚信等情况进行展示,但由于评价指标众多且取值不一,浏览用户无法对商家形成一个综合的信誉评分,也难以对多个商家的信誉进行比较。这里,可以借助因子分析方法来简化指标体系,并通过因子分析得到的各因子方差贡献来构建综合的信誉评分模型。

首先将多个有相互联系的变量表达成少数几个独立的因子。每个因子的重要程度(赋权)是根据其对原始变量总方差的解释贡献来确定的,以此贡献为权重所构建的信誉评价模型,能够比较客观和综合地反映商家信誉的各个方面。针对信誉评价指标体系建立的因子分析模型如下:

$$\begin{cases} x_1 = \mu_1 + a_{11}f_1 + a_{12}f_2 + \cdots + a_{1m}f_m + \varepsilon_1, \\ x_2 = \mu_2 + a_{21}f_1 + a_{22}f_2 + \cdots + a_{2m}f_m + \varepsilon_2, \\ \vdots \\ x_p = \mu_p + a_{p1}f_1 + a_{p2}f_2 + \cdots + a_{pm}f_m + \varepsilon_p. \end{cases}$$

上式中, x_1, x_2, \dots, x_p 为商家信誉评价体系中的各个指标变量,这里分别代表卖家信用度、宝贝描述相符度、卖家服务态度、卖家发货速度、平均退款速度、近30天退款率、近30天投诉率、好评比率和差评比率等。由于商家在近30天处罚数上的差异不大,因此可以忽略该变量^①。 f_1, \dots, f_m 为提取出来的公共因子; $\varepsilon_1, \dots, \varepsilon_p$ 为特殊因子。

由于不同行业的商家在信誉评价指标上的得分差异有很大不同,因此分行业进行因子

① 考虑到各变量的量纲不同,我们使用变量的相对数来进行分析,也即商家在每个变量上的取值都是其与该行业平均水平的相对差异。

分析得到的信誉评分模型会更加准确. 通过软件抓取工具, 我们一共抓取了淘宝网站上 14 个行业共计 39 651 家卖家的信誉评价数据, 然后采用统计软件 SPSS 17.0 对数据进行因子分析. 因子分析结果显示, 所有行业的评价指标最后都可以由四个独立的公因子来表达. 对于“3C 数码”行业, 因子 1 主要与宝贝相符度、服务态度、发货速度有关, 所以可以命名为商家交易态度因子; 因子 2 大致与好评率、差评率、退款速度有关, 因此可以解释为商家售后服务因子; 因子 3 与退款比率、投诉比率高度相关, 因此可命名为商家诚信因子; 最后, 只有信用度因子归属于因子 4, 由于信誉度主要是用来反映商家被好评的交易次数的, 所以因子 4 可以解释为商家交易规模因子. 对于不同的行业, 归属于因子 3 和因子 4 的变量略微有所不同, 变量划分不同的原因主要与不同行业的商品在网上交易的特性有关. 比如“3C 数码”商品的图片和实物感知有很大差别, 因此商家在退款比率和投诉比率这两项指标上的差异会比较大, 而且这两项指标的相关度比较高; 而对于“服饰箱包”类商品来说, 商家在投诉比率上的差异明显较小, 商家的信誉差异更多地体现在销售规模上.

我们可以以各因子的方差解释贡献为权重来构建广告主的综合信誉评分模型, 比如“3C 数码”行业的信誉评分模型为

$$F_{\text{综}} = 0.419f_1 + 0.278f_2 + 0.155f_3 + 0.148f_4. \quad (1)$$

对于每一个广告主的信誉评价指标值, 可以先通过因子分析得到的因子得分系数矩阵将其转化为在各因子上的得分, 然后通过信誉评分模型计算出该广告主的综合信誉值.

4 广告投放最优机制设计

针对某个广告位, 广告商需要在众多的竞争广告中选择一个最合适的广告进行投放, 决策的依据有二: 一是使得所投放的广告能获得用户最大可能的点击; 二是能实现自身收益的最大化. 也就是, 广告商需要在综合考虑点击率和自身收益的情况下选择一个最优的广告, 当然还要考虑广告主的竞价参与约束和他们的竞价策略, 这类似于一个最优机制设计问题. 下面在机制设计的框架下建立广告商的决策模型.

假设有 n 个广告主对这个广告位感兴趣, $N = \{1, \dots, n\}$ 表示这些广告主的集合. 对于每个广告主 i , 我们用 v_i 表示当用户每点击其广告时给他带来的期望收益(以下简称每点击价值), 用 λ_i 表示当他的广告被投放时用户点击其广告的概率(简称点击率), r_i 表示其广告与网页浏览用户偏好的匹配度, ρ_i 表示该广告主的综合信誉度.

对以上参数需要作出一些假定: 对每个广告主来说用户的每点击价值 $v_i (i = 1, \dots, n)$ 是其私人信息, 广告商及其他广告主只知道关于 v_i 的分布, 假设 v_i 在区间 $[\underline{v}_i, \bar{v}_i]$ 上服从分布 $\Phi_i(\cdot)$, 其密度函数为 $\phi_i(\cdot)$, 另外, 广告主之间的点击价值分布 $\Phi_i(\cdot)$ 是相互独立的^①. 所投放广告的点击率不仅依赖于广告和用户偏好之间的匹配度, 还与广告主的信誉度有关. 为此, 假定点击率 λ_i 具有函数形式: $\lambda_i = \rho_i^\alpha \cdot r_i^{1-\alpha} (0 < \alpha < 1)$.

为了竞争得到广告位, 广告主们就广告位的用户每点击价值展开竞价, 不妨用 b_i 表示广告主 i 对用户每点击价值的报价. 针对广告主的竞价向量 $b = (b_1, \dots, b_n)$, 广告商需要决定对广告位的分配和支付. 假设 $\pi_i(b)$ 表示将广告位分配给广告主 i 的概率, $p_i(b)$ 表示当广告主 i 获得广告位时的每点击支付, 那么广告商对广告位的分配规则和支付规则可以分别

① 这是机制设计理论中经典的独立私有估值假设.

表示为 $\pi(b) = (\pi_i(b))_{i \in N}$ 和 $P(b) = (p_i(b))_{i \in N}$.

在此假定下,广告主 i 通过竞价所能获得的收益为: $\lambda_i \pi_i(b)(v_i - p_i(b))$. 对广告商来说,广告主们按照自己的真实点击价值进行报价是最有利的. 为了激励广告主们真实报价 (Telling-Truth), 广告商在设计机制时需要考虑如下的约束条件:

$$\begin{aligned} U_i(v_i) &\equiv E_{v_{-i}} [\lambda_i \pi_i(v_i, v_{-i})(v_i - p_i(v_i, v_{-i}))] \\ &\geq E_{v_{-i}} [\lambda_i \pi_i(b_i, v_{-i})(v_i - p_i(b_i, v_{-i}))], \forall b_i \in [\underline{v}_i, \bar{v}_i]. \end{aligned} \quad (2)$$

也就是,当其他广告主都按自己的真实点击价值 v_{-i} 报价时,广告主 i 也按自己的真实价值报价是其最优的策略,即真实报价是广告主之间竞价的 Nash 均衡,这个约束条件被称为贝叶斯激励相容 (Bayesian Incentive Compatibility) 约束, $E_{v_{-i}}$ 表示对 v_{-i} 的条件期望.

另外,投放机制的设计还要考虑广告主的竞价参与约束. 这里,简单的假定广告主参与竞价的条件是其期望收益非负,即:

$$U_i(v_i) \geq 0, \quad (3)$$

这被称为个人理性 (Individual Rationality) 约束.

广告商就是要在满足以上两个约束的条件下设计一个广告投放机制 (π, P) 使得自身期望收益最大,即:

$$\max_{\{\pi, P\}} \sum_{i=1}^n E_{v_i} \{ E_{v_{-i}} [\lambda_i \pi_i(v_i, v_{-i}) p_i(v_i, v_{-i})] \}. \quad (4)$$

通过简单的变量代换可以在线性框架下讨论上述问题,令 $\vartheta_i(v_i) = E_{v_{-i}} [\lambda_i \pi_i(v_i, v_{-i})]$, $t_i(v_i) = E_{v_{-i}} [\lambda_i \pi_i(v_i, v_{-i}) p_i(v_i, v_{-i})]$, 则 $U_i(v_i) = v_i \vartheta_i(v_i) - t_i(v_i)$.

类似于 Garg 和 Narahari^[14] 等人对关键字广告拍卖机制设计的分析过程,上述的投放机制设计问题可以等价于如下的线性优化问题:

$$\begin{aligned} \max_{\{\pi, P\}} \sum_{i=1}^n \int_{\underline{v}_i}^{\bar{v}_i} [v_i \vartheta_i(v_i) - U_i(v_i)] \phi_i(v_i) dv_i \\ \text{s. t. 对 } \forall i \in N \text{ 和 } v_{-i} \in \prod_{j \neq i} [\underline{v}_j, \bar{v}_j]. \end{aligned} \quad (5)$$

- i) $\vartheta_i(\cdot)$ 关于 v_i 是单调非减的;
- ii) $\pi_i(v_i, v_{-i}) \in [0, 1]$, $\sum_{i=1}^n \pi_i(v_i, v_{-i}) \leq 1$;
- iii) $U_i(v_i) = U_i(\underline{v}_i) + \int_{\underline{v}_i}^{v_i} \vartheta_i(s) ds$;
- iv) $U_i(v_i) \geq 0$.

对以上优化问题求解,得到最优的分配规则 π^* 和支付规则 P^* 分别为:

$$\pi_i^*(v_i, v_{-i}) = \begin{cases} 0 & \text{if } J_i(v_i) < 0 \\ 1 & \text{if } J_i(v_i) = J^{(1)}. \\ 0 & \text{if } J_i(v_i) < J^{(1)} \end{cases} \quad (6)$$

其中, $J_i(v_i) = \lambda_i \left[v_i - \frac{1 - \Phi_i(v_i)}{\phi_i(v_i)} \right]$, $J^{(1)}$ 是 $J_i(v_i)$ ($i = 1, \dots, n$) 中的最大值;

$$p_i^*(v_i, v_{-i}) = \begin{cases} 0, & \text{if } J_i(v_i) < 0 \\ \inf \{ v_i \mid J_i(v_i) > 0 \text{ 且 } J_i(v_i) > J^{(1)} \}, & \text{if } J_i(v_i) = J^{(1)}. \\ 0, & \text{if } J_i(v_i) < J^{(1)} \end{cases} \quad (7)$$

其中 $J_j^{(-i)}$ 表示除 i 以外的 $J_j(v_j)$ ($j \in N, j \neq i$) 中的最大值.

5 广告投放效率测度及仿真分析

广告投放效率是衡量广告投放机制优劣的重要标准. 利用上面设计的机制, 进一步探讨广告投放效率的测度.

5.1 广告投放效率的测度

广告投放效率, 是指就用户搜索目标而言投放广告为满足其偏好所提供的便利程度. 这可以反映在两方面: 一方面, 由于广告为用户浏览和搜索提供的便利程度越高, 用户为辨识目标信息所额外花费的资源即信息代价越少(这不包含处理目标信息本身所花费的资源), 因此可以把用户点击广告所花费的信息代价(即信息转移代价, Information Transition Cost)作为衡量广告投放效率的指标; 另一方面, 所投放广告满足用户偏好的程度越高其期望点击率越高, 从而广告收益越高, 因而也可以用广告投放的总收益来衡量广告投放的效率.

(1) 基于信息代价的测度

根据信息距离(Distance of Information, DIT)理论^[18], 用户点击某一广告位的信息代价(单位为 bit * dit):

$$\lambda \log_2 \frac{1}{\lambda}, \quad \lambda = \rho^a \cdot r^{1-a}. \quad (8)$$

其中 λ 为用户对广告位的潜在点击率. 式(8)符合信息代价的一般定义^[17], 即当 λ 接近于 1 或 0 时用户很容易决定点击与否. 因而对应的信息代价小, 而当 λ 处于 $[0, 1]$ 中间区域时, 用户为辨识广告信息与搜索目标之间的相符度所花费的资源即信息代价会较大. 另外, 广告位的点击率与所投放的广告有关, 因此, 我们这里用 n 个广告主的期望信息代价来作为对广告投放效率的测度指标:

$$\Gamma = \sum_{i=1}^n \pi_i \lambda_i \log_2 \frac{1}{\lambda_i}. \quad (9)$$

其中 π_i 是第 4 节中对广告主 i 的分配概率.

(2) 基于广告收益的测度

广告投放所产生的收益为所有广告主和广告商的收益之和, 根据第 4 节中所述机制, 可知广告投放的总收益为:

$$\Gamma = \sum_{i=1}^n \lambda_i \pi_i (v_i - p_i) + \sum_{i=1}^n \lambda_i \pi_i p_i = \sum_{i=1}^n \lambda_i \pi_i v_i, \quad (10)$$

即为广告位的期望点击价值.

5.2 仿真实验

(1) 广告投放效率的可视分析

对广告投放机制进行可视化分析, 是为了找出广告的投放效率与广告的三个属性即匹配度、信誉度和点击价值之间的隐藏关系. 要可视分析这三个维度间的关系, 需要在二维空间内表现三维信息, 这里采用类似 Radviz 的方法^[18]来建立广告属性分布图.

首先定义广告匹配度、信誉度和点击价值为显示图中三个角的属性指代, 越靠近某角点表示广告的该属性值越大. 由于数据项涉及三个维度, 所以所有数据点都位于三角形中, 三

个指标的相对好坏可以通过其所在坐标位置判断,距离某个顶点越近相应的指标相对越好,图中点的大小对应广告投放效率的高低,点的颜色代表根据用户偏好划分的用户类型.以手机广告为例,这里红色对应“数码发烧友用户”,绿色对应“时尚型手机用户”,蓝色对应“学生手机用户”.在本文机制下,广告投放效率与广告三维属性之间的关系如图 2 所示.

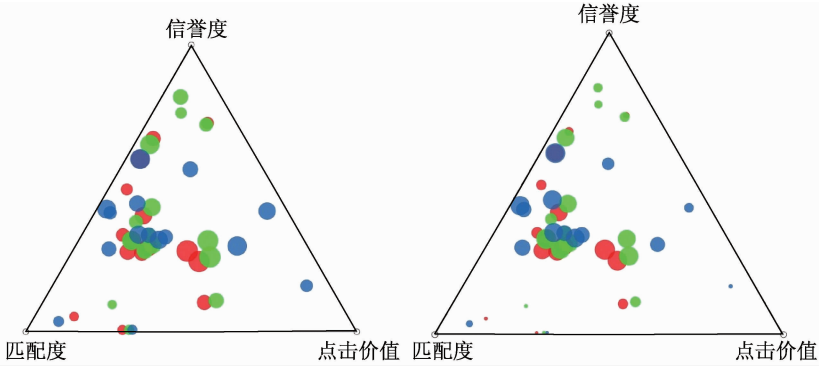


图 2 广告投放效率与广告三维属性之间的关系

Fig. 2 The relation between the quality and attributes of online advantaging

由图 2 可以看出,广告投放效率高的广告一般位于三角形的重心位置,而位于三个顶点区域的广告的投放效率一般都比较低,这说明只有投放那些在匹配度、信誉度和点击价值三个方面综合来看比较好的广告才能获得好的投放效果.目前业界采用的投放策略仅考虑了广告的唯一属性,这不利于广告效益的提升.而且本文的投放机制考虑了信誉因素,大大降低了一些企图通过高竞价来获得广告位的假冒伪劣广告的投放机会.

(2) 信誉对投放效率影响的仿真分析

下面通过仿真实验来说明信誉因素对投放效率的影响,假设有 20 个广告主竞争某一网页广告位,每个广告主的参数确定方法如下:广告主的点击价值 v 随机产生于均值为 1.95 的指数分布,分布和参数的确定依据参见文献[20];匹配度因子 r 与点击价值 v 无关,随机产生于对数均值为 0、标准差为 0.5 的对数正态分布;信誉度因子 ρ 与 r 、 v 都无关,随机产生于对数均值为 0 的对数正态分布; α 取值 1/2.为了说明信誉因素对投放效率的影响,将 ρ 的标准差从 0.005 升至 1(间隔为 0.005),对每个特定的 ρ 标准差随机抽样 500 个广告主点击价值样本和 500 个对应的匹配度和信誉度因子样本,每个样本是 1×20 维的随机向量.

容易验证,在 v 服从指数分布的条件下, $J(v)$ 关于 v 是单调非减的.因此本文对广告位的分配规则简化为:将广告主按 $\lambda \cdot v$ 的降序排列,排名第一的广告被投放.为了说明信誉对投放效率的影响,这里我们与不考虑信誉因素的投放机制作比较.不考虑信誉的投放机制为:仅考虑各广告主的匹配度 r 和点击价值 v ,即按 $r \cdot v$ 的降序排列,排名第一的广告被投放.在两种投放机制下,广告投放效率的均值随 ρ 标准差的变化情况如图 3 所示,(a)图和(b)图分别说明了两种测度方法下的对比情况.

从图 3 可以看出,当广告主之间的信誉度差别不是很大,如 ρ 标准差小于 0.3 时,投放效率几乎不受投放机制的影响.原因是当信誉度差别很小时,按 $\lambda \cdot v$ 排名接近于按 $r \cdot v$ 排名,支付也相近.然而,当信誉度存在较大差别时,不管是按信息代价的测度还是按广告收益的测度,考虑信誉的投放机制都要比不考虑信誉的投放机制的投放效率高,并且二者的差别

随 ρ 标准差的增大而增大. 这是因为, 广告位的点击率与广告主的信誉有关, 当在投放策略中考虑广告主的信誉因素时会有助于提高广告投放效果.

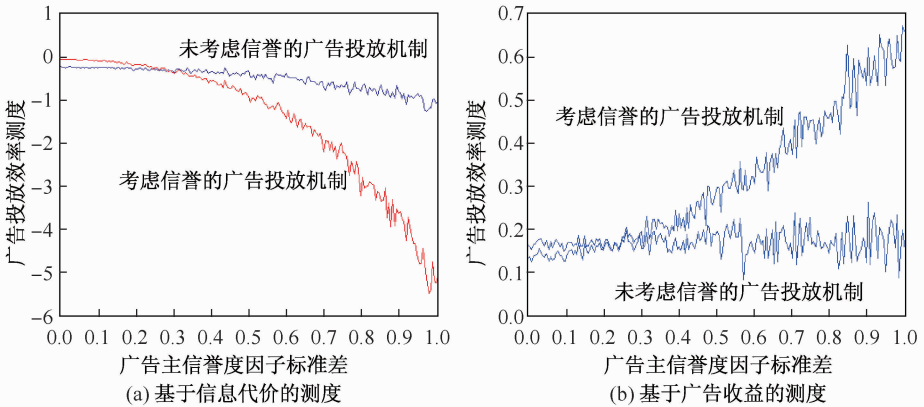


图3 信誉对投放效率影响的仿真实验

Fig. 3 The simulation results of the influence reputation for the quality of online advertising

6 总结与展望

本文研究了考虑信誉的广告投放机制设计和投放效率的测度, 并根据用户行为偏好与广告商品属性的匹配度, 采用因子分析法给出了广告主综合信誉评分方法; 在此基础上运用机制设计理论, 在考虑广告主的竞价参与约束和他们的竞价策略的条件下, 研究了能实现广告收益整体最优的投放机制; 最后通过对广告投放效率测度的仿真实验和可视分析, 证明本文考虑广告主信誉的最优投放机制在广告投放效率方面具有一定的优越性.

由于商品的种类很多, 我们对用户行为的偏好与商品属性的匹配度分析还不够精细, 这还有赖于进一步考虑用户的行为特征进行更精确的定向. 由于机制的设计和计算还比较复杂, 下一步可以考虑在线的计算并能实时地给出最优的广告投放策略. 同时对于广告投放效率的检验仅是基于仿真分析和可视化分析, 如何通过实证分析来验证本文结论这也是我们下一步的工作.

[参 考 文 献]

- [1] AGGARWAL C C, WOLF J L, YU P S. A Framework for the Optimizing of WWW Advertising[M]. Trends in Distributed Systems for Electronic Commerce, LNCS 1402, 1998: 1-10.
- [2] 罗枫. 论网络媒介细分与网络广告投放的人性化[J]. 中国西部科技, 2004(5): 23-24.
- [3] 孙涛. 网络广告系统的用户行为定向研究[D]. 上海: 复旦大学, 2007.
- [4] ADOMAVICIUS G, TUZHILIN A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions[J]. IEEE Trans on Knowledge and Data Eng, 2005, 17(6): 734-749.
- [5] 蔡伟杰, 张晓晖, 朱建秋等. 关联规则挖掘综述[J]. 计算机工程, 2001, 27(5): 31-33.
- [6] LEE J, JUN C H, LEE J, et al. Classification-based collaborative filtering using market basket data[J]. Expert Systems with Applications, 2005, 29(3): 700-704.
- [7] PAPAGELIS M, PLEXOUSAKIS D. Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents[J]. Engineering Applications of Artificial Intelligence, 2005, 18(7): 781-789.
- [8] HAN J, PEI J, YIN Y. Mining frequent patterns without candidate generation[C]//ACM SIGMOD-International

Conference on Management of Data, 2000, Dallas, TX.

- [9] 朱艳春,刘鲁,张巍. 在线信誉系统中的信任模型构建研究[J]. 控制与决策, 2007, 22(4): 413-417.
- [10] 李维安,吴德胜,徐皓. 网上交易中的声誉机制——来自淘宝网的证据[J]. 南开管理评论, 2007, 10(5): 36-46.
- [11] 张仙锋. 信誉的价值:基于淘宝数据对我国电子市场的特色解释[J]. 当代经济科学, 2009, 31(3): 30-39.
- [12] EDELMAN B, OSTROVSKY M, SCHWARZ M. Internet advertising and the generalized second price auction: selling billions of dollars worth of keywords[J]. American Economic Review, 2007, 97(1): 242-259.
- [13] VARIAN H R. Position auctions[J]. International Journal of Industrial Organization, 2007, 25: 1163-1178.
- [14] GARG D, NARAHARI Y, REDDY S S. Design of an Optimal Auction for Sponsored Search Auction[C]//9th IEEE International Conference on E-Commerce Technology and Enterprise Computing, E-Commerce and E-Services (CEC-EEE 2007), 2007: 439-442.
- [15] 卫强,阮楠,单艺. 网络展示广告位置对点击率影响的实证研究[J]. 信息系统学报, 2010, 4(1): 43-52.
- [16] 童强. 搜索引擎关键字广告点击率与保留价研究[D]. 大连:大连理工大学, 2011.
- [17] 姜晖,王浣尘,高朝伟. 基于DIT理论建模的付费搜索排名质量测度研究[J]. 图书情报工作, 2010, 54(2): 113-118.
- [18] PILLAT R M, VALIATI E R A, FREITAS C M D S. Experimental study on evaluation of multidimensional information visualization techniques[C]. American Conference on Human-Computer Interaction, CLIHC, 2005.
- [19] 刘刚,刘万军,张伟. 基于星形用户社区模型的Twitter广告投放[J]. 计算机应用与软件, 2012, 29(4): 44-48.
- [20] 姜晖,王浣尘,关树永. 付费搜索拍卖建模与两类排名机制比较研究[J]. 中国管理科学, 2009, 17(3): 142-149.

(上接第59页)

[参 考 文 献]

- [1] 艾瑞咨询. 2012Q1 中国网络广告市场规模 140.6 亿 网络媒体蓄势待发[EB/OL]. 2012[2013-03-23]. <http://www.iresearch.com.cn/Report/View.aspx?Newsid=170801>.
- [2] BRODE A. Computational advertising[C]//ACM-SIAM Symposium on Discrete Algorithms. Pennsylvania; Society for Industrial and Applied Mathematics Philadelphia, 2008: 992.
- [3] MOBASHER B, DAI H. Improving the effectiveness of collaborative filtering on anonymous web usage data[C]. Proceedings of the IJCAI 2001 Workshop on Intelligent Techniques for Web Personalization (ITWI~I). Berlin: Springer-Verlag, 2001: 49-56.
- [4] SALTON G and LESK M E. Computer evaluation of indexing and text processing [J]. Journal of the ACM, 1968, 15(1): 8-36.
- [5] CHEN K J, LIU S H. Word identification for Mandarin Chinese sentences[C]//Proceedings of the 15th International Conference on Computational Linguistics. Nantes; COLING-92, 1992.
- [6] 陈水利,李敬功,王向公. 模糊集理论及其应用[M]. 第1版. 北京:科学出版社, 2005.