基于改进粒子群算法的网络广告配置优化

陈 亮[1], 张墨华[2]

(1. 河南工业大学 信息科学与工程学院,郑州 450001;2. 河南财经政法大学 计算机与信息工程学院,郑州 450000)

摘 要:针对网络广告的特点,提出了一个基于混合定价策略的网络广告资源配置优化模型,将其建模为一个约束优化问题,最大化网站的总收益。通过罚函数法进行约束处理,提出一种改进的粒子群算法进行求解。仿真结果表明了该算法的有效性。

关键词:网络广告;约束优化;粒子群算法;罚函数;混合定价

中图分类号:TP18;TP393 **文献标志码**:A **文章编号**:1001-3695(2012)01-0142-03

doi:10.3969/j.issn.1001-3695.2012.01.040

Web advertising configuration algorithm based on particle swarm optimization

CHEN Liang[1], ZHANG Mo-hua[2]

(1. *School of Information Science & Technology*, *Henan University of Technology*, *Zhengzhou* 450001, *China*; 2. *College of Computer & Information Engineering*, *Henan University of Economics & Law*, *Zhengzhou* 450000, *China*)

Abstract: According to the properties of Web advertising, this paper proposed a hybrid pricing strategy based model. It was modeled as a constrained optimization problem to maximize total revenue of the Web. Used penalty function method for constraint handling, and proposed an improved particle swarm algorithm for solving the Web advertising optimization problem. The simulation results show the effectiveness of the algorithm.

Key words: Web advertising; constrained optimization; particle swarm algorithm; penalty function; hybrid pricing

随着电子商务的快速发展,网络广告作为一种新的企业营销手段也日显重要。网络广告具有成本低廉、即时互动、效果易评估等特点。从实际来看,目前大多数门户网站的收入均来自网站的广告。因此,网站运营商面临如下具体问题:如何优化配置有限的广告空间,以更好地发挥网络广告的作用,吸引广告客户,从而给自己带来最大收益。文献[1]最早提出广告资源优化模型,此模型以最大化广告点击次数总和为目标;文献[2]对网页属性进行了聚类分析,以减小模型的规模;文献[3]引入熵模型,均衡结果为零的最优变量;文献[4]提出了具有分散惩罚项的目标函数,并用微粒群算法实现。所有这些研究[1~4]都是用点击率来衡量广告效果。实际上网络广告客户希望即使在不点击广告的情况下,网络广告也能影响尽可能多的消费者[5~7]。基于此,文献[8]提出了改进的智能算法进行求解。而这些文献并没有考虑网站和客户的收益问题,即并没有与网站所采用的定价模型联系起来。本文在考虑网站点击率的前提下,给出了基于混合定价策略的网站广告资源优化模型,并采用改进的粒子群算法进行求解。

1 问题描述和建模

本文针对目前网络广告中最主要的广告形式——旗帜广告来研究。所谓旗帜广告(banner)是一个表现商家广告内容的图片,它是目前网络广告中最主要的广告形式[6]。而在网站上,称放置广告图像的网页的每一次显示为一次印象[5,6]。如果客户对广告感兴趣,则会点击广告希望得到更详细的信息,因此,点击次数可以比较客观而准确地反映出广告的效果[7]。把广告的点击次数与广告的印象次数之比称为点击率。显然,点击率是网络广告吸引力的一个重要标志。

网络广告中可用的定价模型有:

a) CPM 定价模型(cost per thousand impressions),即每千人印象成本,是指广告产生每 1 000 个广告印象数的费用[4,6,7]。在 CPM 模型中,广告客户基于广告的印象次数付费给网站,因为不管广告的效果如何,网站都能得到每次印象的收入。因此,这种定价模型更受到网站的欢迎,而广告客户却要承担所有的风险。由于现在的商业竞争非常激烈,这种模型并不合适。

b) CPC 定价模型(cost per thousand click),即千人点击成本,也就是广告图片被点击 1 000 次为基准的网络广告收费模式[5]。CPC 定价更倾向于广告客户的利益。但是,广告的点击率通常是非常低的,因此该定价模型会给网站带来更多风险。

c) 为平衡网站和广告客户间的利益,便出现了混合定价模型。所谓混合定价模型,是 CPM 和 CPC 定价模型的合并[7]。

本文针对旗帜广告进行研究采用的混合定价模型是关于广告的印象次数和点击次数的函数,即印象费用和点击分红之和。本文所考虑的问题是如何优化配置有限广告空间,使网站的收入最大化。

假设某网站受理了 n 个广告业务,记为 $A_i(i=1,2,\cdots,n)$,网站的所有网页按属性分类可分为 m 类($j=1,2,\cdots,m$)(例如新浪网的网页可划分为新闻、体育、娱乐、旅游等不同种

收稿日期:2011-07-03;**修回日期**:2011-08-20

作者简介:陈亮(1978-),男,河南开封人,讲师,主要研究方向为人工智能、网络技术等(sx_wcy@tom.com);张墨华(1979-),男,河南固始人,讲师,主要研究方向为智能信息处理、网络安全等.

类)。假设 x_{ij} 为广告 i 在第 j 类网页上的印象次数(单位为千次印象);g_{ij} 为广告 i 在第 j 类网页上的期望利润;q_{ij} 为广告 i 在第 j 类网页上显示时,每被点击千次时所增加的分红;c_{ij} 为广告 i 在 j 类网页上的点击率;p_{ij} 为广告 i 在第 j 类网页上显示千次时,广告客户付给网站的印象成本,即 CPM;C_i 为第 i 个广告客户的广告预算;M_i 为第 i 个广告客户对广告 i 的最小印象次数的要求(显示率);S_j 为第 j 类网页的容量。

根据上面的描述,则此问题的数学优化模型为

$$f(x)=\max\sum_{i=1}^{n}\sum_{j=1}^{m}x_{ij}(g_{ij}+c_{ij}q_{ij}) \tag{1}$$

$$\text{subject to}\ \sum_{j=1}^{m}x_{ij}(p_{ij}+c_{ij}q_{ij})\leqslant C_i \tag{2}$$

$$\sum_{j=1}^{m}x_{ij}\geqslant M_i \tag{3}$$

$$\sum_{j=1}^{m}x_{ij}\leqslant S_i \tag{4}$$

其中:式(1)为网站收入最大的目标函数,约束条件式(2)是指每个广告的支出不能超过其预算(广告的印象费用和点击分红之和);约束条件式(3)是广告客户所要求达到的最小广告印象次数;约束条件式(4)是指所投放的广告不能超过某一类网页的容量。本文优化目标与以往模型不同之处在于引入了混合定价模型,建立了在一定广告预算下,网站总体收入最大化的决策模型。

2　基于粒子群算法的网络广告资源优化求解过程

2.1　免疫粒子群算法概述

粒子群算法是一种仿生算法,模仿鸟类的觅食行为,将问题的搜索空间类比于鸟类的飞行空间,将每只鸟抽象为一个无质量无体积的粒子,用以表征问题的一个候选解,优化所需要寻找的最优解则等同于要寻找的食物[9,10]。粒子群算法具有记忆粒子最佳位置的能力和粒子间信息共享的机制。一般,粒子群算法采用简单的速度—位移模型,核心要素是一群粒子在搜索空间中的位置(表示决策变量)和运动速度。设第 i 个粒子的位置的 $x_i=(x_{i1},x_{i2},\cdots,x_{in})$ 为,它经历的最好位置为 p_{best},粒子群体所经历过的最好位置为 g_{best},第 i 个粒子的速度表示为 $v_i=(v_{i1},v_{i2},\cdots,v_{in})$。对每一次迭代,粒子的速度和位置则根据如下方程变化:

$$v_i^{k+1}=\omega v_i^k+c_1r_1(p_{\text{best}}^k(i)-x_i^k)+c_2r_2(g_{\text{best}}^k-x_i^k) \tag{5}$$

$$x_i^{k+1}=x_i^k+v_i^{k+1} \tag{6}$$

其中:c_1、c_2 为学习因子,分别调节全局和个体最好粒子飞行方向的最大步长,通常情况令 $c_1=c_2=2$;r_1、r_2 为在(0,1)内变化的随机数;w 为惯性权重。粒子群算法的特点是收敛速度快、易于实现,但多样性较差,使得后期收敛速度明显变慢[10],易陷入局部最优解。

免疫优化算法主要模拟生物免疫系统的信息处理机制,如学习、记忆、调节等功能[9,10],已经在工程应用领域显示出了强大的应用潜力。基于此,本文将免疫算法与粒子群优化相结合,在保留粒子群算法优点的前提下,引入免疫信息处理机制,将免疫算法中免疫记忆功能与自我调节机制引入粒子群算法,保持粒子维持一定的浓度,以保持种群的多样性。

第 i 个粒子的浓度定义如下:

$$D(x_i)=\frac{1}{\sum_{j=1}^{n}|f(x_i)-f(x_j)|};\ j=1,2,\cdots,n$$

其中:$f(\cdot)$ 为相应的适应度值,$|\cdot|$ 为两者的欧式距离。

基于上述粒子浓度的概率选择公式定义为

$$P(x_i)=\frac{\dfrac{1}{D(x_i)}}{\sum_{i=1}^{n}\dfrac{1}{D(x_i)}}=\frac{\sum_{j=1}^{n}|f(x_i)-f(x_j)|}{\sum_{i=1}^{n}\sum_{j=1}^{n}|f(x_i)-f(x_j)|} \tag{7}$$

由式(7)可知,与抗体 i 相似的抗体越多,抗体 i 被选中的概率就越小。这样使得低适应度个体也可以获得进化的机会。因此基于浓度的概率选择公式理论上保证了抗体的多样性。

2.2　算法具体求解过程

1)初始化过程　由上面的分析可知,用矩阵表达决策变量 $x_{ij}(i=1,2,\cdots,n;j=1,2,\cdots,m)$,则

$$X=\begin{bmatrix}x_{11},x_{12},\cdots,x_{1m}\\x_{21},x_{22},\cdots,x_{2m}\\\cdots\\x_{n1},x_{n2},\cdots,x_{nm}\end{bmatrix}$$

在本算法中,粒子坐标用 $X=\{x_{ij}\}$ 表示,描述粒子在 $m\times n$ 维空间中的运动位置,而粒子的运动速度由矩阵 $V=\{v_{ij}\}$ 表示。

2)适应度函数　网络广告资源优化是约束优化问题,这里通过罚函数进行无约束化处理,各个约束处理过程如下:

$$g_1(x)=C_i-\sum_{j=1}^{m}x_{ij}(p_{ij}+c_{ij}q_{ij})\geqslant 0$$

$$g_2(x)=\sum_{j=1}^{m}x_{ij}-M_i\geqslant 0$$

$$g_3(x)=\sum_{j=1}^{m}x_{ij}-S_i\geqslant 0$$

这里,定义适应度函数为

$$F(x)=f(x)-a_1\max(0,g_1(x))-a_2\max(0,g_2(x))-a_3\max(0,g_3(x)) \tag{8}$$

其中:a_1、a_2、a_3 为惩罚因子。

2.3　算法具体步骤

a)确定免疫粒子群算法的参数值,包括学习因子 c_1、c_2、最大迭代次数 $k_{\max}$、惯性权重 ω,群体规模 N。

b)运用初始化过程,随机生成 K 个粒子 x_i 和初始化速度 v_i,形成初始化粒子群 p_0。

c)根据适应度函数式(8),计算每个粒子适应度值,找到粒子的个体极值 $p_{\text{best}}(i)$ 和全局极值 g_{best}。

d)按照式(5)和(6)更新粒子的速度和位置,并将 g_{best} 对应的位置粒子存入记忆库。

e)根据基于浓度的粒子选择概率式(7),从 N 个粒子中依次选取 K 个粒子。

f)以记忆库中的粒子代替适应度最差的粒子,新生成一个粒子群 p_1,进入下一次迭代。

g)根据最大迭代次数判断是否结束迭代,并输出符合条件的最优粒子;否则转 c)。

3　仿真实验与结果

在 MATLAB 7.0 下对算法进行了实现,并对多个问题进行了仿真分析。这里以 20 个广告客户,10 类属性页的情况进行说明。每组数据运算 50 次。由于网站的统计数据涉及商业机密很难获得,在这里仿照文献[3,8],在实验的环境下产生点击率的方法来生成不同广告在不同类属性页上的点击率,并与文献[8]保持一致。其中,参数设置如下:x_{ij} 的范围为[0,200],g_{ij} 为[20,800]间均匀分布的随机数,$q_{ij}=10\times p_{ij}$,$g_{ij}=0.4\times p_{ij}$,M 取[70,400]之间均匀分布的随机数,C 取[500000,

2000000]之间均匀分布的随机数,S 取[2000,4500]间均匀分布的随机数。

免疫粒子群算法的参数设置为:最大迭代次数 $k_{\max}=1\,000$,种群大小 $N=100$, 学习因子 $c_1=c_2=2$,惯性权重 $\omega=1$,惩罚因子参数 a_1、a_2、a_3 分别为2、1、0.5。

经过运算,算法得到的最优解,即每个广告在各类网页上印象次数如表1所示,表明了算法的有效性。

表1　算法最优解

j	i																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	30	41	12	76	79	75	7	126	103	52	104	49	95	112	78	8	135	11	59	90
2	96	40	56	48	159	140	23	69	18	128	25	75	94	3	58	119	27	116	48	40
3	83	9	116	136	7	60	0	69	143	112	45	34	3	98	39	39	109	79	78	6
4	39	38	133	7	0	87	52	152	69	33	10	85	110	82	70	0	87	75	47	13
5	52	114	158	16	0	113	173	6	129	8	12	96	23	37	49	120	77	63	106	10
6	87	17	79	40	76	108	76	82	37	66	66	14	10	91	171	47	77	40	107	46
7	137	46	34	171	168	87	156	103	13	3	69	119	101	118	89	98	93	8	163	88
8	58	60	71	89	0	48	8	59	109	8	59	66	73	55	13	132	2	39	5	119
9	90	93	76	6	82	85	184	0	49	62	75	69	34	135	76	39	78	3	99	89
10	8	53	108	26	28	79	39	78	13	47	48	109	49	124	0	67	97	29	55	12

从表1中可以看到,有的印象次数比较少,说明广告和此类网页的相关性很小,所以点击率偏低。因此,广告在这类网页上投放的次数相对就要少些,否则无法得到很好的广告效果。反之,如果相关性很高,则点击率也就很高,那么在这类网页上可以相对多投放一些广告。

为了说明本文算法的效果,将本文算法与文献[8]算法进行了对比分析,得到的最终优化结果(网站最大化收益)如图1所示。从图1的结果可知,本文算法由于设计了适合问题表示的相关算子,加快了收敛速度,求解效率较高;本文算法得到的最优结果优于文献[8],说明本文算法采取的各种策略是有效的,可以得到更好的优化结果。算法运行结束,本文算法得到的网站最大收入为438.8万元,高于文献[8]算法的422.5万元。

4　结束语

本文提出了一种基于混合定价模型的网络广告资源优化模型,平衡了网站广告客户和网站运营商的利益,更适合实际应用情况。此模型是一个约束优化问题,通过罚函数将其转换为无约束问题,并通过改进粒子群算法进行求解。仿真实验结果表明了算法的有效性。

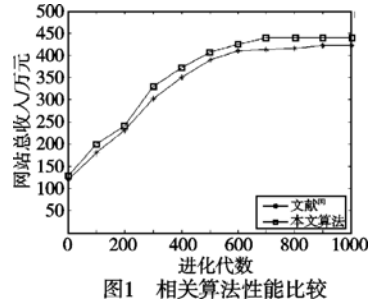

图1　相关算法性能比较

参考文献:

[1] LANGHEINRICH M, NAKAMURA A, ABE N, *et al*. Unintrusive customization techniques for web advertising[J]. Computer Networks, 2003, 31(11-16): 1259-1272.

[2] NAKAMURA A, ABE N. Improvements to the linear programming based scheduling of Web advertisements[J]. Electronic Commerce Research, 2008, 5(1): 75-98.

[3] BOSKAMP V, KNOOPS A, FRASINCAR F, *et al*. Maximizing revenue with allocation of multiple advertisements on a Web banner[J]. Computers and Operations Research, 2011, 38(10): 1412-1424.

[4] 齐洁,汪定伟. 求解网络广告资源优化模型的改进微粒群算法[J]. 控制与决策, 2004, 19(8): 881-884.

[5] 李震. 网络广告效果的综合评价研究[J]. 科技管理研究, 2010, 12(8): 321-325.

[6] 王建冬. 网络广告界面评价中的人机交互理论——网络广告交互界面与广告效果关系模型[J]. 现代图书情报技术, 2009, 3(10): 301-304.

[7] 张志华,王莉. 网络环境下广告资源优化决策模型[J]. 鞍山科技大学学报, 2006, 31(2): 321-315.

[8] ZHAO Wen-dan, WANG Ding-wei. Modeling and algorithm of Web advertising resources optimization[C]//Proc of the 29th Chinese Control Conference. 2010: 5590-5594.

[9] 刘朝华,张英杰,章兢. 一种双态免疫微粒群算法[J]. 控制理论与应用, 2011, 28(1): 65-72.

[10] 刘朝华,张英杰,章兢,等. 基于免疫双态微粒群的混沌系统自抗扰控制[J]. 物理学报, 2011, 60(1): 019501.

(上接第141页)　表3　两种算法计算结果比较

算法	$d_{\min}$	$d_{\max}$	d_{mean}	w/%	t_{mean}
本文算法	0	1	1.00	96	3.87
基本 PSO	0	5	2.25	84	3.63

从表3可以看出,本文算法所得到调度方案的精确度比基本PSO算法方案高。由于本文算法在利用粒子群算法求得优化解后,还要利用拟牛顿方法进行局部优化,以尽量达到或接近全局最优点,所以其运算时间比基本PSO算法要稍长。总体来说,本文算法在缩短项目工期、合理调配项目资源上有比较好的效果,其求解质量要优于基本PSO算法。

4　结束语

本文针对以项目工期最小为目标的资源受限项目调度问题,设计了一种混合粒子群算法,并与基本PSO算法在解决资源受限项目调度问题时的精确度和效率进行了比较。计算测试表明,本文算法解决资源约束项目调度问题相对于基本PSO算法有较大优势,这说明本文方案有效。

参考文献:

[1] KOLISCH R, HARTMANN S. Experimental investigation of heuristics for resource-constrained project scheduling: an update [J]. European Journal of Operational Research, 2006, 174(1): 23-37.

[2] 杜焱,彭武良. 一种求解资源受限项目调度问题的遗传算法[J]. 沈阳理工大学学报, 2009, 28(1): 70-74.

[3] 郑超,高连生. 蚁群算法在资源受限项目调度问题中的应用[J]. 计算机工程与应用, 2005, 41(27): 205-208.

[4] 喻小光,战德臣,聂兰顺. 应用遗传模拟退火算法实现资源受限项目调度[J]. 计算机工程与应用, 2009, 45(24): 17-20.

[5] 王巍, 赵国杰. 粒子群优化在资源受限工程调度问题中的应用[J]. 哈尔滨工业大学学报, 2007, 39(4): 669-772.

[6] 袁亚湘,孙文瑜. 最优化理论与方法[M]. 北京: 科学出版社, 2009.

[7] KOLISCH R. Serial and parallel resource-constrained project scheduling methods revisited: theory and computation [J]. European Journal of Operational Research, 2009, 99(2): 320-333.