

Dr. Dingxi Qiu is an Assistant Professor of Industrial Engineering at University of Miami, whose research interests in direct marketing are mainly on statistical modeling and data analysis. He can be reached at dingxi@miami.edu

Abstract

Internet advertisement via web banner ads has become popular in recent years. The advertiser usually has a large set of candidate host websites to choose from to put banner ads on for marketing campaigns. We develop a model to select host websites for Internet campaigns and to allocate budgets properly to each selected website so as to maximize the overall expected profit from showing banner ads to different types of Internet online visitors.

Keywords and Phrases: Banner ad; Cost; Revenue; Decaying effect

1. Introduction

Internet marketing has become one of the most important channels of advertising in the last 10-15 years (Bakos, 2001; Manchanda, et al., 2006). Many companies not only display their products and services at their home website, but also purchase advertisement spaces from popular host websites (Hofacker and Murphy, 2000). Host websites can cover a variety of potential customers, such as news readers in different regions and people in different clubs. These host websites can often help the companies to publicize their products and services by attracting the attention of online visitors with well-designed banner ads. Some visitors may choose to click through the banner ads and be directed to the advertiser's website that carries out the business, hence bringing in profit. The total amount of money that the host website charges the advertiser for showing banner ads varies, depending on the popularity of the host website and the magnitude of inventory of advertisement space; while the benefit of purchasing the banner ad impressions can be expressed as the banner ad's lifetime value of capturing a new customer from the web

surfing interface. Different websites charge differently, but the success of advertisement might solely depends on the number of impressions and the online visitors' specific attributes, including the web-specific attributes, such as the type of website, and the visitor-specific attributes, such as the number of clicks the online visitors have made.

The marketing company usually has a large pool of candidate host websites to choose for Internet marketing campaigns. Deciding which host websites to post their advertisements on and optimally allocating their budget for each host website at the beginning of each decision period have long been a difficult question (Jain and Singh, 2002). Qiu and Malthouse (2009) have shown that people who have never seen the banner ads are, on expectation, more profitable than people who have already been contacted for a long time, but did not make any purchases or transactions. Hence, exposing banner ads to diverse groups of online visitors can help improve the expected revenue from the marketing campaign. Existing data mining techniques can help decide which websites provide relatively mutually exclusive customer sets. Properly allocating the budget to these host websites can provide the marketing company competitive advantage. In this paper, we develop a method to help marketing company choose host websites from a large candidate set of host websites and decide the proper decision rules, regarding when to show banner ad to a particular online visitor, for each chosen host website. We assume that each host website has a fixed activation cost for each Internet advertisement campaign period. The host website also charges a constant variable cost per impression after the activation fee is paid. The variable cost per impression for Internet advertisement is usually very low and the host website often charges the advertisers on a per-thousand-impression basis or per-ten-thousand-impression basis. We also assume risk-neutral utility of expected revenues and hence aim to maximize the total expected revenue from the marketing campaign. Before we discuss the issues related to multiple host websites, let us discuss the decision making process on a single host website.

2. Decisions on a Single Host Website

Both the banner ad exposure information and click information can be stored in the computer cookie files and sometimes also in the host website's remote servers. When the web-surfer requests the download of a webpage, the host website or the advertiser has a fraction of second to decide whether to show banner ad to this particular online visitor. The information stored in the cookie file can help evaluate the profitability of showing the banner ad.

Potential customers sometimes do not react immediately to the marketing campaign, and it

often takes multiple impressions to acquire a single customer. Each impression can be considered as having a lifetime value that decays as a function of time. Qiu and Malthouse (2009) proposed a model that considers the lifetime value of each impression so that the decision regarding showing a banner ad depends on the number of impressions that the online visitors have received, among other important factors. Let us assume that we can extract p measurements for each online visitor, $\mathbf{x} = \{x_1, x_2, \dots, x_p\}$. These measurements include time-dependent variables, such as number of impressions within the last two days, and static random variables, such as user internet connection speed. Her/his enrollment probability can then be predicted by using discrete survival analysis model. Technically we divide the total exposure period into mutually exclusive time segments and fit a logit model for the success probability by considering all the segments. The probability of successful acquisition can be transformed into a nonlinear function of \mathbf{x} and t , where t represents time since the first impression of banner ads for this online visitor.

$$P(\mathbf{x}, t) = \frac{\exp\{\alpha_0 + \alpha_1 \log(t) + \beta_1 x_1 + \dots + \beta_p x_p\}}{1 + \exp\{\alpha_0 + \alpha_1 \log(t) + \beta_1 x_1 + \dots + \beta_p x_p\}}.$$

The host website shows the banner ad to the online visitor if and only if the expected benefit, $r = P(\mathbf{x}, t) \times (\text{expected revenue per captured customer})$, of showing the banner ad is higher than a predetermined threshold value, say τ . In other words, the host website only attracts online visitors whose expected revenue is higher than this threshold cost.

The threshold τ is a decision variable in the model for the host website and it determines the total number of impressions the advertiser buys for the decision period. Intuitively, we can rank the online visitors according to their profitability. Let $f(r)$ be expected number of online visitors to this website in terms of a function of r . Then maximizing the revenue from advertisement campaign is equivalent to the following optimization problem:

$$\max \int_{\tau}^{\infty} r f(r) dr - a - b \int_{\tau}^{\infty} f(r) dr \quad (2.1)$$

$$s.t. \quad a + b \int_{\tau}^{\infty} f(r) dr \leq B \quad (2.2)$$

$$\int_{\tau}^{\infty} r f(r) dr - a - b \int_{\tau}^{\infty} f(r) dr \geq 0, \quad (2.3)$$

where a is the fixed activation cost and b is the variable cost and B is the maximum allowable budget. Constraints (2.2) and (2.3) provide lower bound and upper bound for τ respectively. This constrained optimization problem (2.1) is one-dimensional in nature and hence can be easily solved

with any commercial optimization software package, provided that $f(r)$ can be accurately evaluated based on the available history data.

3. Decision on Multiple Host Websites

If multiple host websites are available for advertisement, then the advertiser usually needs to select a subset of host websites for their Internet marketing campaign. The advertiser also needs to allocate the budget in each selected website wisely so as to maximize the total expected profit. These two questions are, in fact, intertwined, because the threshold for each host website is a function of the budget for any pre-determined $f(r)$, which in turn affects the total revenue. Let $W = \{w_1, w_2, \dots, w_n\}$ be a set of n available host websites that can carry out business for the marketing company. Let the constant cost and variable cost of showing banner ads on those websites be a_i and b_i ($i = 1, \dots, n$), respectively. Also let $f_i(r)$ be the expected number of online visitors at host website i . The objective of the marketing campaign is then to maximize the overall benefit of online advertisement on all the selected host websites,

$$\max_{W_D \subset W, \tau_i} \sum_{i=1}^n I\{w_i \in W_D\} \left[\int_{\tau_i}^{\infty} r f_i(r) dr - a_i - b_i \int_{\tau_i}^{\infty} f(r) dr \right], \quad (3.1)$$

where $W_D \subset W$ is the chosen subset of W for advertisement and τ_i are the threshold variable for the revenue on host website i , regarding whether to show a banner ad to a online visitor. Both W_D and τ_i (where $w_i \in W_D$) are decision variables in the objective function. The optimization problem (3.1) is an NP-hard problem since the total number of candidate sets W_D increases exponentially as n increases; however, if for some reason, we can finalize the subset of candidates for marketing campaign, then the following proposition can help make decisions on the optimal allocation of the budgets.

Proposition 1 *Assume that the first m websites have been chosen for advertisement in one particular period. The threshold τ_i ($i = 1, \dots, m$) on different host websites should be proportional to its variable cost, and the following equilibrium equation should always be satisfied,*

$$\frac{\tau_1}{b_1} = \frac{\tau_2}{b_2} = \dots = \frac{\tau_m}{b_m} = \tau_0. \quad (3.2)$$

Proof: Maximizing the objective function (3.1) is equivalent to

$$\max_{\tau_1, \dots, \tau_m} \sum_{i=1}^m \int_{\tau_i}^{\infty} (r - b_i) f_i(r) dr$$

because the candidate websites have been selected and therefore their fixed costs $\sum_{i=1}^m a_i$ on these m websites will be encountered anyway. Intuitively, the online visitors to each candidate website that, on expectation, can bring high profit to the advertiser are preferred for impression. If online visitors can be chosen in a sequential way, then the last online visitor selected for online campaign should be the least profitable one. Let ROI denote return on investment of the marketing budget on the last online visitor, then the ROI based on variable cost only will be $ROI_i = (\tau_i - b_i)/b_i$. Suppose that equation(3.2) does not hold, then there must exist two websites, i and j , such that $\tau_i/b_i > \tau_j/b_j$. This is equivalent to $ROI_i > ROI_j$. This obviously cannot be true because if this is the case, we can move some of the budget from host website j to website i to improve the overall objective function (3.1). ■

In this paper, we employ a greedy algorithm to solve (3.1). Intuitively, there should be an optimal number and a feasible region for the number of impressions for advertisement. If too few impressions are sent out, the average cost will be too high; on the other hand, if too many impressions are sent out, many not-so-profitable online visitors will be included for advertisement. In our method, we first find out the most profitable website and then use boundary constraints on τ_i to eliminate those whose feasible solution regions do not have overlap with that of the most profitable website. If the total number of remaining host websites is small after this elimination process, we can enumerate all possible subsets $W_D \subset W$ and optimize the objective function for each W_D . The solution corresponding to the highest objective value will be considered as the best solution. The optimization of (3.1) for each pre-specified W_D is done in one-dimensional space with τ_0 as the single decision variable.

4. Empirical example

An American Internet Service Provider (ISP) advertises its products and services in more than one hundred websites in its history. Due to the economical constraints, they have to advertise in a limited number of websites in the following campaign period. The ISP chose the following twenty candidate websites for investigation because they represent the most popular websites with diverse online visitor sets. The diversity of the websites can be verified by comparing the so-called *pairwise*

We chose the fifth website *doubleclick.com* as an example and the revenue and cost comparison for that website is shown in Figure 1. The x-axis shows the number of impressions in a unit of 10,000 impressions. We assume that doubleclick.com charges a fixed cost of US\$1,000.00 and a variable cost of US\$ 1.00 per 10,000 impressions. As we can see from the figure that the gross revenue is lower than the average cost if we show less than 985,000 impressions from this website. In addition, the net revenue is a concave function of the number of impressions.

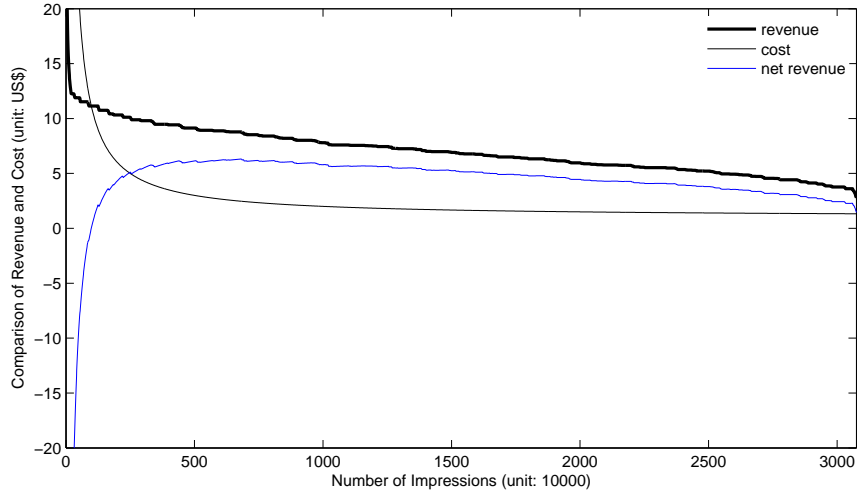


Figure 1: Comparison of Revenue and Cost for Banner Ads posted on doubleclick.com

Suppose our budget in the following month is US\$70,000. Also assume that all websites charge the same activation cost and variable cost. The objective of this study is to identify the most profitable host website portfolio for advertisement. We first decrease the candidate host website set size by comparing the threshold boundaries. Suppose we can finalize our choice to three websites doubleclick.com, espn.go.com and cbs.com. Then, we can enumerate all $2^3 - 1 = 7$ possible candidate subsets W_D and optimize (3.1) in terms of τ_0 with a precision of 75. One candidate solution (but not necessarily optimal) to this model is obtained manually with the numbers of impressions for the three websites $z_5 = 2750$, $z_7 = 750$ and $z_{11} = 500$, and with identical thresholds $\tau_5 = \tau_7 = \tau_{11} = 3.0$. The corresponding overall objective function value is approximately 18125.

5. Conclusion and Discussions

This paper proposed a method to help advertisers to optimize the expected profit by choosing a set of host websites for marketing campaign. We showed that the thresholds for revenue on each

website should satisfy the equilibrium equation. The overall profit can be optimized in greedy fashion and hence is easily implementable. We presented an empirical example from a ISP provider and use the real data to show how to use the proposed optimization strategy.

This method can be extended to optimize advertisement efforts in other channels such as TV commercials. The advertisement contents in TV channel can be optimized by using the user's personal information in the cable company under the user's written permission. The customized campaign may impose additional computation and managerial efforts for the company; hence the operational cost is higher in general. However, in the long run, this methodology will be beneficial to both the advertisers and the cable companies. This strategy should also be welcome by TV watchers because most people prefer to see things that are interesting to them during the advertisement breaks, rather than seeing anything irrelevant.

Acknowledgments: Dingxi's portion of this research was financially supported by the Summer Award in Natural Sciences and Engineering at University of Miami in 2008.

REFERENCES

1. Bakos, Y.(2001). The emerging landscape for retail E-commerce, *The Journal of Economic Perspectives*, 15(1), 69-80.
2. Hofacker, C.F. and Murphy, J. (2000). Clickable World Wide Web banner ads and content sites, *Journal of Interactive Marketing*, 14(1), 49-59.
3. Jain, D., and Singh S. S. (2002). Customer lifetime value research in marketing: A review and future directions, *Journal of Interactive Marketing*, 16(2), 34-46.
4. Manchanda, P., Dubé, J. P., Goh, K. Y., and Chintagunta, P.K. (2006). The effect of banner advertising on Internet purchasing, *Journal of Marketing Research*, 43, 98-108.
5. Qiu, D. and Malthouse E. C. (2009). Quantifying the indirect effects of a marketing contact, *Expert Systems with Applications*, 36, 6446-6452.