Expert Systems with Applications 38 (2011) 4234–4243

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

How to measure the effectiveness of online advertising in online marketplaces

Cookhwan Kim [a], Kwiseok Kwon [b], Woojin Chang [a,*]

[a] Department of Industrial Engineering, Seoul National University, 599, Kwanak Street, Kwanak-Gu, Seoul, Republic of Korea
[b] Department of e-Business, Anyang Technical College, Anyang, Kyeonggi, Republic of Korea

ARTICLE INFO

Keywords:
Online marketplace
Online advertising
Advertising effect model
Poisson-gamma model
Click-through data
Hierarchical Bayes model

ABSTRACT

The online marketplace, in the form of an "open market" where a very large number of buyers and sellers participate, has occupied a rapidly increasing position in e-commerce, resulting in sellers' increasing investment in online advertising. Hence, there is a growing need to identify the effectiveness of online advertising in online marketplaces such as eBay.com. However, it is problematic to directly apply the existing online advertising effect models to the click-through data of online marketplaces. Therefore, a model must be developed to estimate the effectiveness of online advertising in the online marketplace in terms of its characteristics. In this paper, we develop an analytical Bayesian approach to modeling click-through data by employing the Poisson-gamma distribution. Our results have implications for online advertising effect measurement, and may help guide advertisers in decision-making.

1. Introduction

Online advertising (ad) is a form of promotion that uses the Internet and World Wide Web for the express purpose of delivering marketing messages to attract customers. The online advertising industry is expected to be stable and manifest a continuing upward trend until 2011. The compound annual growth rate is anticipated to increase by 17.4% during this period (2007–2011) and touch the $197.11 billion mark. In the coming years, online advertising spending is expected to overtake the TV advertising market. The rapid growth of this industry is being driven by the increasing number of Internet users, rising awareness, and growing broadband subscription rate and e-commerce, which is playing a key role in this industry.

Not surprisingly, predicting the effectiveness of online advertising has gained much research attention. Since the Internet opened up to the general public in the mid-1990s, a database consisting of repeated customer visits to a Web site along with individual advertising exposures can be obtained by cookies to track users. Many studies (e.g., Chatterjee, Hoffman, & Novak, 2003; Chan Yun, Kihan, & Stout, 2004; Manchanda, Jp, Goh, & Chintagunta, 2006) have researched online advertising effect models by making use of the data. It became a new fashion to study at the individual level the analysis of advertising effects. A traditional sales-advertising model and banner advertising model has been developed and applied to estimate the online advertising effect.

Meanwhile, the growing prevalence of Internet access has enabled new markets to emerge online. An e-marketplace is an electronic exchange where firms or individuals register as sellers or buyers to communicate and conduct business over the Internet. There are many types of e-marketplaces based on a range of business models such as business-to-business (B2B), business-to-consumer (B2C), or consumer-to-consumer (C2C). Perhaps the best-known marketplace except for B2B is eBay.com, an enormous globally available Auction house for products. We call a marketplace such as eBay an "online marketplace." Thereby, the online marketplace has occupied a rapid increasing position in e-commerce. In Korea, where the B2C e-commerce market amounts to $12 billion, online marketplaces, including Auction (http://www.auction.co.kr, an eBay company in Korea), formed approximately 50% of the market in 2008.

Online marketplaces have taken the form of an "open market" where buyers and sellers are easily registered and conduct business. This has resulted in a very large number of buyers and sellers in the marketplace and increased competition among sellers. Hence, the sellers in the marketplace have increasingly invested in online advertising there so that their products are widely exposed to buyers. Hence, there is a growing need to identify the effectiveness of online advertising in the online marketplaces to help the sellers running their business in the marketplaces.

However, applying the existing online advertising effect models to estimate the effectiveness of online advertising using click-through data in a specific online marketplace can create problems. Since customers in an online marketplace have an intention to purchase, they are exposed to the ads relatively quickly, and customers decide on purchases within a very limited number of visits compared to customers on other Web sites such as Web portals.

* Corresponding author. Tel.: +82 2 880 8335; fax: +82 2 889 8560.
E-mail address: changw@snu.ac.tr (C. Kim).

doi:10.1016/j.eswa.2010.09.090

Moreover, the previous models are not appropriate for utilizing click-through data gathered in online marketplaces. Banner advertising models cannot incorporate click-through data from customers who rarely visit a Web site before they drop out. Traditional sales-advertising models cannot be applied to count data generated by click-throughs either because the model was originally designed to provide the relationship between advertising exposures and sales.

Therefore, a model must be developed to estimate the effectiveness of online advertising in the online marketplace in terms of its characteristics.

The current article is designed to respond to the aforementioned research need. Specifically, the primary objective of this study is to develop a Bayesian model to understand the effectiveness of online advertising in the online marketplace; and to evaluate the feasibility of the proposed model using actual consumer behavior data in an online marketplace.

2. Theoretical background

2.1. Sales-advertising model

Online advertising through the Internet draws attention to the contrast between traditional assumptions about advertising and its effects and the realities in the current online marketplace. Most of the models that reflect traditional assumptions for the carry-over effects of advertising are designed in such a way that the sales (S) as a response value would be affected by the lagged values of the advertising variables (A). This sales response function is represented by a multiplicative functional form to permit diminishing returns to scale, and a log–log transformation makes this relationship linear (Parsons, 1976):

$$\ln S_t = \alpha + \beta \ln A_t + \gamma \ln A_{t-1} + \theta \ln A_{t-2} + \cdots \tag{1}$$

Parsons (1976) took the simplest finite horizon version of the model, which involves only current advertising and advertising in the most recent previous period:

$$\ln S_t = \alpha + \beta \ln A_t + \gamma \ln A_{t-1}. \tag{2}$$

This concept of "distributed lags" was first used and discussed by Irving Fisher in 1925. In addition, the Dutch econometrician Koyck published his monograph *Distributed Lags and Investment Analysis* in 1954; thus, the use of distributed lags became widespread in work of an econometric nature. The simplest case is that the effect of the independent variables upon the dependent variable starts declining in constant proportion from the first period on:

$$\begin{aligned} S_t &= \alpha + \beta_1 A_t + \beta_2 A_{t-1} + \beta_3 A_{t-2} + \cdots + u_t, \\ &\quad \textit{where } \beta_1 = a;\ \beta_2 = a\lambda;\ \beta_3 = a\lambda^2;\ 0 < \lambda < 1. \end{aligned} \tag{3}$$

It can be rewritten as the following form:

$$S_t = (1 - \lambda)\alpha + aA_t + \lambda S_{t-1} + u_t - \lambda u_{t-1}. \tag{4}$$

The prior models that embody the concept of distributed lags use a considerable number of lagged exogenous variables, while the simple Koyck model uses only one lagged and one non-lagged exogenous variable (Palda, 1965). Bass and Clarke (1972) showed six possible alternative models to the Koyck model, each of which is an extension of the Koyck model to a second or higher-order lag function. Moreover, A_t is lagged by one or more periods in addition to the lag in the dependent variable. The following equation is an example of the extension:

$$\begin{aligned} S_t &= (1 - \lambda_1 - \lambda_2)\alpha + \beta a_0 A_t + \beta a_1 A_{t-1} + \lambda_1 S_{t-1} + \lambda_2 S_{t-2} \\ &\quad + u_t - \lambda_1 u_{t-1} - \lambda_2 u_{t-2}. \end{aligned} \tag{5}$$

Givon and Horsky (1990) developed a model that combines advertising retention over time and purchase feedback across competing brands. The authors added two things to the Koyck model: (1) the Markovian transition matrix corresponding to individuals' brand-switching behavior among brands A and B (where brand B may represent all non-A brands) and (2) the relative price of brand A at time t. Leone (1983) built a model that could solve the problems involving the presence of autocorrelation, multicollinearity, or high seasonality among given competing brands in data by using multivariate time series analysis. His model is also based on the Koyck distributed lag form that implies a geometric decay of advertising. In 1981, the consumer model was proposed to figure out the micro-model of the aggregate sales-advertising relationship for a single product. This model incorporates two factors that cannot be seen in other models: reach of the ads and the rate of decay of their effectiveness over time. Blattberg and Jeuland (1981) assumed an exponentially decaying effectiveness function to measure the carry-over effects of an ad because the consumer gradually forgets the advertisement.

2.2. Advertising effect model for repeated visits

Up until the late 1990s, the majority of the papers with "measuring advertising effects" in the title were written to figure out the direct relationship between advertising exposure and sales. Since the Internet was opened up to the general public in the mid-1990s, a database consisting of repeated customer visits to a Web site along with individual advertising exposures has been obtained by cookies to track users. Many studies are based on a large sample of real data from an online banner advertising company because banner advertising is most commonly sold on a cost-per-mille (CPM) or click-through-rate (CTR) basis. Nowadays, banner advertisement provides an interactive advertising platform to convey the advertising message to the target audience. Most banner advertisements are colorful or animated to catch the customer's eye instead of a passive billboard just sitting in the background hoping to be read. Chan Yun et al. (2004) examined the effects of animated banner ads, as well as the moderating effects of involvement, on each stage of the hierarchy of effects model, and explored the applicability of the hierarchy of effects model to the banner advertising environment through an online experiment. The results from this study demonstrate that animated banner ads prompt better advertising effects than static ads. The effects were measured by click-through rates. The click-through rate is also significant for analyzing the advertising effects of a static banner. However, estimating the probability of a purchase or a revisit is more important than the interactivities between advertisements and customers. Manchanda et al. (2006) specified an individual-level joint purchase timing and expenditure model as a function of advertising exposure and found that exposure to banner advertising has a significant effect on Internet purchasing behavior. They modeled an increase in purchase probability as a function of banner advertising exposure via a semi-parametric survival model. Chatterjee et al. (2003) focused on repetition effects in advertising and modeled advertising response based on content-laden Web sites with embedded banner advertising. Related prior research suggests two different patterns of consumer response to repeated advertising exposures within the same Web session. The first pattern posits that response probability decreases over time (Buchanan & Morrison, 1985). The second response pattern holds that initial response probability may be low, but increases with repetition to a maximum level and then diminishes over subsequent repetitions (Berlyne, 1970). Chatterjee et al. (2003) theorized that wearout dominates in online advertising environments so that, for most consumers, there are relatively strong diminishing returns to early repeated exposures that taper off as exposures continue. The logit regression model is natural to use for fitting binary response variables such as purchase or revisit. Montgomery,