Expert Systems with Applications 38 (2011) 9867–9878

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Random effects model for estimating effectiveness of advertising in online marketplaces

Cookhwan Kim [a], Sungsik Park [a], Yongseok Chang [b], Woojin Chang [a,*]

[a] Department of Industrial Engineering, Seoul National University, 599, Kwanak Street, Kwanak-Gu, Seoul, Republic of Korea
[b] Investment Banking I, Korea Development Bank Capital, KDB Capital Building, 16 Yeouido-dong, Youngdeungpo-gu, Seoul, Republic of Korea

ARTICLE INFO

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

ABSTRACT

This paper presents an application of the Bayesian Markov Chain Monte Carlo (MCMC) used to select cost-effective ad spots in online marketplaces. Due to the rise of electronic commerce, the online advertising industry which is highly complex undergoes rapid changes. And there are plenty of studies that keep coming up with the similar methodologies to predict click-through rates for ad spots. Previous research has mainly considered the following models: a logistic regression model and a binomial model connected by a linear link function. However, it is problematic to directly apply the existing online advertising effect models to the click-through data of online marketplaces. Because generally a click-through rate is fairly low so that a small change in its rate might give a somewhat larger prediction error in terms of click-throughs. We propose a Bayesian Poisson-gamma model to predict click-throughs instead of their rates and further extend to incorporate random effects in order to account for heterogeneity of variance between keywords. Our results may help guide online advertisers in decision-making.

1. Introduction

As the Internet usage continues to surge, the online advertising industry grows dramatically. For instance, the volume of the online advertising market in South Korea skyrocketed from $6.625 billion in 2005 to $12.409 billion in 2007 while the volume of whole advertising market only increases from $70.539 billion to $79.897. It means that not only the portion of online advertising market surged but also the importance of the online advertising market soared. In the next few years, the size of the online advertising market can overtake the TV advertising one. Increasing Internet penetration rate and the ones' usage and surging e-commerce transaction are the major causes that lead rapid growth of the online advertising market. In this circumstance, it can be said that now Internet has become the crucial medium for advertising such as traditional medium (e.g., TV, radio, newspaper, magazine).

Recently the support of database allows researcher to analyze the data at the individual level and it become a new fashion of analyzing advertising effect at the personal level. It is no surprise that there has been a lot of research conducted on measuring the effectiveness of online advertising at the personal level (e.g., Chatterjee, Hoffman, & Novak, 2003; Manchanda, Dube, Goh, & Chintagunta, 2006). Especially a traditional sales-advertising model and banner advertising model have been developed and applied to estimate the online advertising effect.

When e-commerce is in its infancy, people are not accustomed to buying the product in online mall and afraid of being deceived especially when they are going to buy products from unknown sellers (e.g., C2C e-market place). However, as the volume of transactions in online market increases, people get used to shop in the online mall and the size of the one gets bigger and bigger. Especially in the Internet, "open market" has emerged as a place where people can register easily and do their business (e.g., eBay), and we name the online mall such as eBay an "online marketplace".

In Korea, buying products in online marketplaces is convenient and quick because approximately half of Korean people live in Seoul and its environs, products can be delivered the next day. Thanks to the convenience and swiftness, the volume of business in Korean online market in 2007 surges to $17.765 billion which is 5 times bigger than the volume in 2001. In this circumstance, analyzing the advertising effect in online marketplace is important and the issue about the online marketplaces' commercial is relatively fresh. Also, advertising types and system in Korea online marketplace are more diverse and prosperous than other countries'. Therefore, it is a meaningful work to measure advertising effects using Korean online marketplaces' data.

However, applying the existing banner advertising model or traditional sales-advertising model to the data without revision is not appropriate because visitors have strong intention to purchase and the number of visits to buy the product is very limited in the online marketplaces. Furthermore, recent studies that are focused

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

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.02.044

on solving the online advertising budget allocation problem adopt optimization methods with assumptions that might not be true in actual cases.

Thus, we suggest a Poisson-gamma model with random effects which is suitable to the consumers' click data in online marketplaces.

2. Literature review

When it comes to research on the sales-advertising relationship, the presence of an advertising budget constraint can be a criterion to classify related research papers into two groups. When the advertising budget is limited, the optimal budget allocation is the main issue for advertisers as well as researchers. Holthausen and Assmus (1982) presented a model for the allocation of an advertising budget to geographic market segments when the sales response to advertising in each segment is characterized by a probability distribution. Gensch and Welam (1973) formulated a model for determining the optimal allocation of a given advertising budget over M interacting market segments and a time domain of T periods. As online advertising has become prevalent as an important marketing tool, some researchers also have paid attention to the budget allocation problem on the Internet. Zhao and Nagurney (2005) addressed the determination and evaluation of optimal internet marketing strategies when a firm can advertise on multiple websites. And the model's objective is finding optimal amount of click-through subject to a budget constraint. Nakamura and Abe (2005) construct a linear program model to maximize the clicks for banner ads by finding the banner ad placement. They investigate that how variety variables affect the response function. However we cannot apply the findings to the niche markets served by keyword advertising.

Freuchter and Dou (2005) address that how to dynamically allocate resources with a given budget when company can advertise on multiple Web portals by using the keyword-activated banner ads on the Internet. And they divide portals into generalized portal and specialized portal. And using techniques of dynamic programming, they find solutions for the optimal budgeting decisions. The finding is that more generalized ads at generic portals yield higher click-through rates.

Without consideration of budget limitation is the other stream of research on the advertising-sales relationship. In this case, the study includes the measuring advertising effect alone. Koyck (1954), Palda (1965) suggest model that measure the cumulative advertising effects using time series model, frequently employed and well known as Koyck model. Bass and Clarke (1972) test six different distributed lag models and suggest statistical models of the dynamics of sales and advertising need not be limited to the highly restrictive Koyck model. Parsons (1976) propose a Rachet model based on sales response function is represented by a multiplicative form rather than linear one in order to permit diminishing returns to scale. Haley (1978) observed that when advertising intensity surge the sales response also surge and exponentially decrease. And Blattberg and Jeuland (1981) suggest the advertising-sales model which assumes exponentially decaying effectiveness.

However since Internet becomes the "new media" for advertising, there has been lots of research on capturing effectiveness of Internet advertising, especially about banner advertising (Chatterjee et al., 2003) and keyword advertising in portals such as Google and Yahoo! (Chen, Liu, & Whinston, 2009).

In this paper we will present a model that captures the effectiveness of advertising in "online marketplaces". It is a relatively new issue which is branched from Internet advertising but not included in banner advertising as shown in Fig. 1.

3. Theoretical background

3.1. 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):

$$\ln S_t = \alpha + \beta \ln A_t + \gamma \ln A_{t-1} + \theta \ln A_{t-2} + \cdots. \quad (1)$$

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

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

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:

$$S_t = \alpha + \beta_1 A_t + \beta_2 A_{t-1} + \beta_3 A_{t-2} + \cdots + u_t,$$
$$\text{where} \quad \beta_1 = a;\ \ \beta_2 = a\lambda;\ \ \beta_3 = a\lambda^2;\ \ 0 < \lambda < 1. \quad (3)$$

It can be rewritten as the following form:

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

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 6 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:

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

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 micromodel 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.