Expert Systems with Applications 39 (2012) 12727–12742

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

An empirical test to forecast the sales rank of a keyword advertisement using a hierarchical Bayes model

Cookhwan Kim [a,*], Sungsik Park [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
Impression
Sales rank
Hierarchical Bayes model

ABSTRACT

Online advertising (ad) is a form of promotion that uses the Internet and World Wide Web for the expressed purpose of delivering marketing messages to attract customers. Not surprisingly, how to predict the effectiveness of online advertising has gained lots of research attention. This study introduces the hierarchical Bayesian analysis to the online advertising effect model involving competition with other products. It developed a competition model with a time-decaying effect that is applicable for the sales-rank data in the online marketplace. The proposed model formalizing the hierarchical structure has performed better than the reduced model without having random effect components. It captures the heterogeneous advertising responses across the products as well as search keywords. Our results have implications for online advertising effect measurement, and may help guide advertisers in decision-making.

1. Introduction

Modern advertising market research involves a wide variety of interrelated and often quite complicated processes, such as competition between different brands of a single product type. Particularly in the online marketplaces, not only competing brands for advertisers' own products but also differentiated products positioned in different branches of the market structure can be listed in a same search result page due to a common keyword in their titles. Even though both similar-brand products and dissimilar-brand products seem to be in the same product category, they might have different levels of consumers' preferences on each product attribute. The primary goal for this research is to forecast the sales rank of each corresponding product, given that a product can appear on multiple pages depending on the stored descriptive keywords in its title. For each page, the number of impressions is daily updated, which represents that consumers' demands for each keyword are changed on a daily basis. Consumers would respond to some of the listed products before click-through by referring the information including a displayed product image, the cumulative number of buyers, the number of reviews, a satisfaction score and ads posted on the top of the page. After clicking on the title of a desired product, they look through the contents of the landing page such as the average delivery time. Consumers' purchases immediately reflect the rank order of products, which is actually based on the cumulative sales occurred within the latest three days. In the next section, we will build a prediction model for the sales rank of each product with regard to the overall situation.

2. Literature reviews on the sales forecasts for effective advertising

A large number of response models have been proposed in the literature linking online advertising to sales or market shares. One of the most popular models in advertising-sales literature is Koyck's model (Koyck, 1954), which was derived from the direct duration interval model (Clarke, 1976). However, despite of its structurally flexibility, Koyck's simple specification of geometrically decaying lagged effects of advertising has been questioned many times for its restrictive hypothesis (Mariel, 2005). Blattberg and Jeuland (1981) have also derived a related model, an aggregate sales-advertising response model by postulating a consumer micro-model. Different from the previous work based on a rigid model of geometrically decaying lagged advertising effects, they adopted an exponentially decaying effectiveness function by assuming that the consumer gradually forgets the advertisement.

Meanwhile, implicitly or explicitly all competing businesses employ a strategy to select a mix of marketing resources (Carpenter, 1987) so that marketing researchers including Eskin, Baron and Wittink (Eskin & Baron, 1977; Wittink, 1977) studied on the interaction effects between advertising and other marketing mix variables such as price. Recent papers extended the literature on advertising response models by considering both interaction

* Corresponding author. Tel.: +82 10 7714 6436; fax: +82 2 889 8564.
E-mail address: cookfan@naver.com (C. Kim).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.01.175

effects and dynamic effects of different themes of advertising because managers, policy makers and researchers would find how to allocate a firm's given advertising budget across different advertising themes to improve sales performance in competitive environment (Naik, 2003). To figure out the dynamic effects in multi-theme advertising, many of them employed the standard dynamic linear model (DLM) as follows.

$$y_t = F_t\Phi_t + \beta X_t + \varepsilon_t, \quad \Phi_t = H_t\Phi_{t-1} + u_t + \omega_t \tag{1}$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ and $\omega_t \sim N(0, W)$

Please see West and Harrison's book for more details of the standard DLM (West & Harrison, 1997). Typically, methods using the Bayesian MCMC provide a rigorous framework for estimating the dynamic system. Neelamegham and Chintagunta (Neelamegham & Chintagunta, 2004) extended the standard DLM by inserting the pooling equation defined in the second stage:

Stage 1. Observation equation

$$y_{bt} = F1_{bt}\Phi1_{bt} + \upsilon1_{bt}, \quad (\upsilon1_{bt}|V1, \Sigma) \sim N(0, V1, \Sigma) \tag{2}$$

The dependent variables are specified as a function of brand model specific (b), time-varying parameters (t). $F1_{bt}$ is a matrix containing explanatory variables associated with each of the variables in y_{bt}. Φ_{bt} consists of time varying brand-model level parameters.

Stage 2. Pooling equation

$$\Phi1_{bt} = F2_t\Phi2_t + \upsilon2_{bt}, \quad (\upsilon2_{bt}|V2, \Sigma) \sim N(0, V2, \Sigma) \tag{3}$$

The second stage specifies separates parameters which vary only over time from ones of the first stage. This hierarchical structure enables to pool information across brand-models. It pools the first stage parameters by specifying a mean parameter vector for $\Phi1_{bt}$. The Eq. (3) implies that $\Phi1_{bt}$ is a linear combination of the mean values across all brand-models, $\Phi2_t$ and the "error" term $\upsilon2_{bt}$. This equation is also known as the "structural" equation.

Stage 3. System equation

$$\Phi2_t = G\Phi2_{t-1} + \omega_t, \quad (\omega_t|W, \Sigma) \sim N(0, W, \Sigma) \tag{4}$$

It is equivalent to the system equation of a dynamic linear model (DLM). The third stage specifies time evolution of the parameters that come from the second stage. $\Phi2_t$ evolves through time and can be characterized by a random walk process because we do not expect drastic jumps in parameters a few periods, forward or backward. The distribution of ω_t is conditional on an error matrix, W and the variance scale matrix, Σ. $V1$ and $V2$ are also variance parameters. In this paper, we would follow an example of a multi-theme advertisement based on the dynamic linear model approach and tried to link marketing mix variables to sales directly. However, due to fluctuating sales demands and other complex variations influencing purchase behaviors, it might be hard to predict sales quantity accurately at each time point. Since total sales for the past three days determine the priority of display, called 'sales rank' of the product list on a search result page, especially in Korean online marketplaces, both variables are likely to be highly correlated. In addition, taking into account the distributed lag period of 1–2 days, at least four variables including $Sales_t$, $Sales_{t-1}$, $Salesrank_t$ and $Salesrank_{t-1}$ jointly affect the structure of the situations which we try to analyze. Even though the structural equation containing a set of interdependent/highly correlated variables results in a high R-squared (for instance, more than 98%) and low standard error (for instance, less than plus or minus 0.05), it is difficult to evaluate the effects of heterogeneous marketing mix variables incorporated with the four variables. Therefore, it would be more reasonable to predict a sales rank instead of a sales quantity for each product because the demand of the whole market tends to be distributed over the top competitors for a corresponding keyword. In this manner, the structural feature of our proposed model is denoted as follows. First, we omit the first stage depicted as the observation equation, Eq. (2), because for that stage we have multiple observation sequences that are highly correlated each other. Second, we set the sales rank at each time point as a scalar dependent variable and employ the system equation of the third stage where the current status of a sequence is dependent upon its previous status. It can be distinguished from the first-order autoregressive model because parameters for estimating the sensitivity of the sales rank are linked to the pooling equation as noted in Eq. (3). Finally, in the pooling equation at the second level hierarchy, the explanatory variables are divided into two groups, product-specific and keyword-specific variables. These statements are formally expressed in Section 2.

3. Model specification

3.1. Basic model

Consider a product i that is sold in an online marketplace. Ideally, our dependent variable would be log of sales of a product on a particular site. Even though a product is sold in multiple websites or physical stores in practice, we consider only a single marketing channel by assuming that online advertisements in a certain online marketplace do not affect the consumer purchase behavior in other online websites or offline stores. Since sales of products vary across product sectors and firm sizes, we would not directly build the model that predicts the values of product sales. Instead, we would like to use the sales rank for a corresponding product as a dependent variable. In Eq. (6), the reason for using log-transformation rather than levels is that the log-transformation measures the effect of a change in the independent variables on the percentage change in the dependent variable. This is a proper step to take because there are scale effects in our case. Exogenously, the number of keyword impressions for a popular product is much larger than that for an unpopular product. The purchase conversion rate per visit is plausibly a function of consumer reviews on the web. Even though the ideal dependent variable in this study is log sales, we use the log rank of product sales. Prior literature has shown that the relationship between sales rank and sales follows a Pareto distribution (Chen, Liu, et al., 2008):

$$sales = \beta_1 rank^{\beta_2} \tag{5}$$

Moreover, Schnapp and Allwine (2001) mapped the relationship between ranks and sales using a sample of books on 'Amazon.com' and found that the relationship between log ranks and log sales is close to linear. Due to the linear relationship between them, the estimated coefficients in our specification and standard errors would be scaled by a constant. In Korean online marketplaces, a sales rank that dynamically changes over time is generally determined by the total amount of sales for the last three days. Therefore, we expect the current sales rank to be correlated with the previous sales rank. The sales rank on the search result page for a specific keyword during time t is a function of the sales rank for the previous period ($rank_{ij}^{t-1}$), a product fixed effect (Ψ_i), and a fixed effect for the search result page of keyword j (μ_j) and other related factors. The product fixed effect is associated with the factors such as the word-of-mouth effect, the quality of the product and the popularity of the brand. The fixed effect for a search result page is related to the preferences of the customers of the search keyword. That is,

$$\begin{aligned}\ln(rank_{ij}^t) = \lambda \ln(rank_{ij}^{t-1}) + \Psi_i + \mu_j + \alpha_j \ln(A_{ik}^t) + \Theta \ln(P_i) \\ + Z\Gamma^t + D\Pi + \varepsilon_{ij}^t\end{aligned} \tag{6}$$

$$A_{ij}^t = q\exp(\phi_1 x_{ij1}^t + \phi_2 x_{ij2}^t)\eta_j^t + q(q-1)e^{-\omega}\exp(\phi_1 x_{ij1}^{t-1} + \phi_2 x_{ij2}^{t-1})\eta_j^{t-1} \tag{7}$$

where the rank denotes the sales rank. Here, we consider ranking all products in reverse order so that rank 1 becomes rank N, rank 2