DOI: 10.7763/IPEDR. 2012. V52. 15

The Model of Purchasing and Visiting Behavior of Customers in an E-commerce Site for Consumers

Shota Sato[1+] and Yumi Asahi [2]

[1] Department of Engineering Management Science, Tokyo University of Science, Japan

[2] Management of Business Development, Shizuoka University, Japan

Abstract. Recently, the e-commerce site market value in Japan is growing. Many companies in the e-commerce market compete for customers with each other. Therefore e-commerce managers need to predict customer's purchasing and visiting behavior. We developed the model which capture customer's purchase and visit behavior in this study. As a result of analysis, purchase and visit tendencies of each customer were revealed. The model of this study will helps e-commerce managers to capture customer's purchase and visit. Our model can be further extended by using a combination data of e-commerce shops and real shops.

Keywords: Cliskstream Data, Hierarchical Bayes Model, Internet

1. Introduction

Recently, the e-commerce site market value in Japan is growing[3]. Many companies will compete in the e-commerce market. Many companies in e-commerce market compete for customers with each other. Therefore, e-commerce managers need to decide which customer is good customer who may benefit in several days. In other words, e-commerce managers need to understand customer's behaviour in an e-commerce site, and need to predict which customer will purchase and visit in several days.

In this study, we focus on customer's purchasing and visiting behaviour. Predicting customer's purchasing helps e-commerce managers to decide which customer is good customer who may benefit. It also helps e-commerce managers to understand when customers intend to purchase. Customer's purchasing in e-commerce sites is one of the major problems for e-commerce managers. Several preceding studies focused on customer's purchasing in e-commerce sites. For example, Moe and Feder[4] developed a model which predicts customer's purchasing . Bucklin and Sismerio[2] also proposed a model to predict purchasing. They decomposed the user's purchasing process into the several steps such as "completion of product configuration". Van den Poel and Buckinx[6] analyzed purchasing behaviour and they found that variables such as "Number of days since last visit" have strong effect on purchasing. These studies computed the probability of purchasing given that a customer visited the e-commerce site. However, it is unclear when the customer visits the e-commerce site. Preceding studies don't consider customer's visit. Capturing both customer's purchasing and visiting helps e-commerce managers to understand when customers intend to purchase and visit the website. Therefore, in this study, we aim to model customer's two decision stages, visiting and purchasing behaviour. As an example of studies on several customer's decision stages, Bucklin and Lattin[1] proposed a model of purchase incident and brand choice. They decomposed the probability of brand choice into two decision stages, brand choice and purchase incident. In this study, probability of purchasing is decomposed into conditional probability of purchasing given a visit and probability of visiting.

Purposes of this study are to develop a model which captures both customer's purchasing and visiting, and to get managerial insights on marketing strategies to customers.

2. Model

2.1. Simultaneous probability model of purchasing and visiting

In this study, a model which focus on customer's purchasing and visiting is developed. Customers don't purchase in the e-commerce site unless they visit the site. A binary variable y_{it} equals one if customer *i*

[+] Corresponding author. Tel.: + 81-3-5228-8351;
E-mail address: s_sato@ms.kagu.tus.ac.jp.

(i=1,…,N) purchased something on the t^{th} day(t=1,…T) and zero no purchase was made. A binary variable v_{it} equals one if customer i visited the site on the t^{th} day and zero if customer i didn't visit the site at the t^{th} day. A binary choice model is employed in this study. Let $P(y_{it}, v_{it})$ be simultaneous probability of purchasing and visiting. Simultaneous probability of purchasing and visiting by customer i on the t^{th} day, $P(y_{it}=1, v_{it}=1)$, is written as

$$P(y_{it}=1, v_{it}=1) = P(y_{it}=1 \mid v_{it}=1)P(v_{it}=1) \tag{1}$$

where $P(y_{it}=1|v_{it}=1)$ is conditional probability of purchasing given customer i visited the site on the t^{th} day and $P(v_{it}=1)$ is visit probability of customer i on the t^{th} day. The probability which customer i visit the site on the t^{th} day but don't purchase anything, $P(y_{it}=0, v_{it}=1)$, is written as

$$\begin{aligned} P(y_{it}=0, v_{it}=1) &= P(y_{it}=0 \mid v_{it}=1)P(v_{it}=1) \\ &= \{1-P(y_{it}=1 \mid v_{it}=1)\}P(v_{it}=1) \end{aligned} \tag{2}$$

The log likelihood function is written as

$$LL = \sum_{i=1}^{N}\sum_{t=1}^{T}\left[y_{it}v_{it}\log\{P(y_{it}=1, v_{it}=1)\} + (1-y_{it})v_{it}\log\{P(y_{it}=0, v_{it}=1)\} + (1-v_{it})\log P(v_{it}=0)\right] \tag{3}$$

Parameters are estimated by using equation (3).

2.2. Modeling purchasing and visiting behavior

A binomial logit model is employed. We model the conditional purchasing probability with the binomial logit as

$$P(y_{it}=1 \mid v_{it}=1) = \frac{\exp(u_{it})}{1+\exp(u_{it})} \tag{4}$$

$$u_{it} = x'_{it}\beta_i \tag{5}$$

where u_{it} is the utility of customer i at the t^{th} day, x_{it} is a vector of independent variables which include an intercept, β_i is parameter vector of customer i. In this study, parameter vector β_i is different for each customer in order to capture heterogeneity.

We model the visit probability with the binomial logit as

$$P(v_{it}=1) = \frac{\exp(z_{it})}{1+\exp(z_{it})} \tag{6}$$

$$z_{it} = w'_{it}\gamma_i \tag{7}$$

where z_{it} is the utility of customer i on the t^{th} day, w_{it} is a vector of independent variables which include an intercept, γ_i is parameter vector of customer i. In this study, parameter vector γ_i is also different for each customer in order to capture heterogeneity.

2.3. Hierarchical model

In order to capture heterogeneity, hierarchical Bayes model is employed. A parameter vector β_i is lumped together with parameter vector γ_i as parameter vector θ_i. In other words, parameter vectors are written as

$$\beta_i = (\beta_{i1}, \dots, \beta_{ik_1})' \tag{8}$$

$$\gamma_i = (\gamma_{i1}, \dots, \gamma_{ik_2})' \tag{9}$$

$$\theta_i = (\beta_{i1}, \dots, \beta_{ik_1}, \gamma_{i1}, \dots, \gamma_{ik_2})' \tag{10}$$

where k_1 is the number of dependent variables including a intercept of purchase and k_2 is the number of dependent variables including a intercept of visit.

Parameter vector θ_i is specified as

$$\theta_i = \Delta + \eta_i \qquad \eta_i \sim MVN(0, V_\beta) \tag{11}$$

where Δ is parameter, and $MVN(0, V_\beta)$ is multivariate normal distribution with a zero mean vector, covariance matrix V_β. Multivariate normal distribution is used as a prior distribution of Δ. Inverse Wishart distribution is used as a prior distribution of V_β. Parameters of prior distributions were chosen as prior distributions are diffuse but proper distributions[5].

3. Data

3.1. The detail of data

Data of Joint Association Study Group of Management Science were used. This data contains clickstream data, transaction data, and customer information data of an e-commerce site which sells golf products in Japan. Data was collected from July 1st 2010 to June 28th 2011. In Japan, a huge earthquake occurred in March 11th 2011. Therefore data from July 1st 2010 to March 10th 2011 is used. Data from July 1st 2010 to January 31th 2011 was used as an estimation period and the remaining period was set aside as a holdout period. Customers are qualified for inclusion in the sample if they purchase something more than 5 days in estimation period. In this study, data were summarized at daily level because we aim to capture customer's purchase and visit behavior each day.

3.2. Independent variables

In this study, two variables are used as independent variables x_{it} and w_{it}. Variables we use in this study are shown in table 1. The variable CumPages is included in the model to capture the relationship between purchasing and visiting behaviour and the number of product pages which a customer viewed in the past. The variable CumMoeny is included in the model to capture the relationship between purchasing and visiting behaviour and the total of money which a customer spent in the past. CumMoney was defined as the 10^{-4} yen. On July 1st , the beginning day of data, one dollar equals around 80 yen. A Square root of CumPages was used as an independent variable in order to reduce the effect of outliers.

Tab. 1: independent variables

The names of variables	description
CumPages	the number of product pages which customer i viewed between the (t-7)th day and the (t-1)th day which a customer don't purchase.
CumMoney	the total of money which customer i spent between the (t-30)th day and the (t-1)th day

4. Empirical analysis

4.1. Estimation method

Parameters are estimated using Markov Chain Monte Carlo (MCMC) method. A Metropolis-Hastings (MH) algorithm is used to draw samples of parameters θ_i and Gibbs sampler is used to draw samples of parameters Δ, V_β. 100000 draws were used for burn in, and additional 100000 draws were used to infer the posterior distribution of the parameters. We kept 5th draw to reduce computer memory requirement. The resulting 20000 draws were used in our analysis. Posterior means are used as estimates in this study.

4.2. Result

First, table 2 shows posterior means of parameters Δ. Parameters Δ indicates the average effect on purchase and visit. The sign "*" in table 2 indicates the 95% confidence interval of the parameter doesn't contain zero. Table 2 shows CumMoney has negative effect on purchase on average. A customer will not purchase after he/she spent much money.

Tab. 2 ： parameters of Δ

Δ					
purchase			visit		
intercept	CumPages	CumMoney	intercept	CumPages	CumMoney
-2.12*	0.05	-0.29*	-1.03*	0.17*	0.05

CumPages has no effect on purchase on average. Most of customers may view product pages and evaluate products on the day they purchase products. On the other hand, table 2 shows CumPages has positive effect on visit on average. This result shows a customer tends to visit the site after s/he viewed many product pages.

Next, figure 1 shows posterior means of parameters β_i for each customer. Parameters β_i means the effect on purchase. Dots in figure 1 represent customers and the number attached to dots represents the serial number of customer. A horizontal axis in figure 1 means the parameter of CumPages for purchasing, a vertical axis in figure 1 means the parameter of CumMoney for puchasing. For example, customers No.4, No. 75 are located in the upper right side in figure 1. Therefore they tend to purchase after they viewed much product pages. Customers No.47, No. 8 are located in the lower right side in figure 1. Therefore they tend to purchase after they viewed many product pages and after they didn't spend much money. They may be good customer who may benefit at the time they viewed many product pages and don't spend money because they may have intention to purchase.

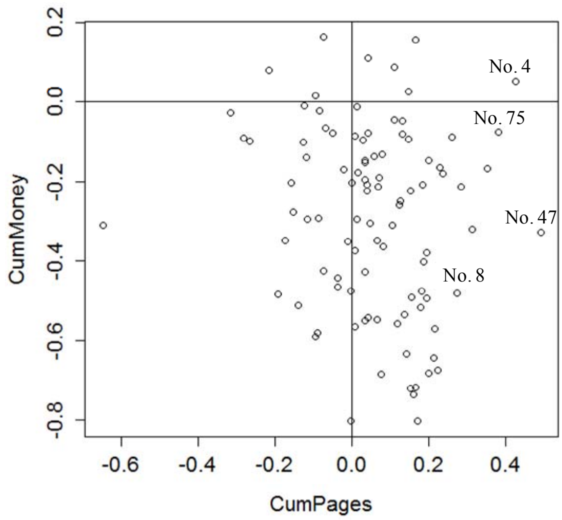

Fig.1: scatter plot of parameters of CumPages and CumMoney

Next, figure 2 shows posterior means of parameters β_i, γ_i for each customer. We focus on the parameter β_i of the variable CumPages and the parameter γ_i of the variable CumPages. A horizontal axis in figure 2 means the parameter of CumPages (purchase) and a vertical axis means the parameter of CumPages (visit). Customer No. 4 and No.47 are located in the lower right side. They tend to purchase after they viewed many product pages though they don't tend to visit after they viewed many product pages. Customer No. 34 is located in the upper left. Customer No. 34 tends to visit after viewing many product pages though customer No. 34 doesn't tend to purchase after viewing many product pages.

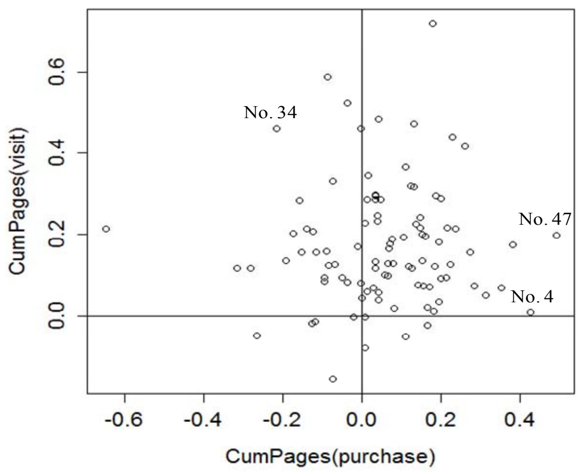

Fig.2:scatter plot of parameters of CumPages(purchase) and CumPages(visit)

Last, the performance of the model is investigated. The performance of the model is investigated by Receiver Operation Characteristics (ROC) curve and Area under the curve (AUC). AUC ranges 0.5 to 1 and is 0.5 in case of predicting randomly. Therefore, if AUC of the model exceed 0.5, it is interpreted that the performance of the model is better than random prediction. Purchasing possibility is calculated by estimates and data of holdout period. We calculate purchase possibility of customer *i* on tth day, P(y_{it} =1, v_{it}=1), as

$$P(y_{it}=1, v_{it}=1) = \frac{\exp(u_{it})}{1+\exp(u_{it})} \frac{\exp(z_{it})}{1+\exp(z_{it})} \tag{12}$$

The performance of the model of estimate period was 0.705 of AUC and that of holdout period was 0.650 of AUC. The result of AUC shows the performance of the model is better than random prediction.

5. Conclusion

Purposes of this study are to develop a model which captures both customer's purchasing and visiting behaviour, and to get managerial insights on marketing strategies to customers. As a result of analysis, it was found that the total of money which a customer spent in the past 30 days has the negative effect on purchasing on the whole. It was also found that the total number of product pages which a customer viewed in the past 7 days has positive effect on visiting on a whole. Moreover, by capturing both purchasing and visiting behaviour, not only the tendency of purchasing of customers but also the tendency of visiting of customers was revealed. As a result of investigating the performance of the model, it was found that it was better than random prediction.

The model of this study will provide useful knowledge to e-commerce managers for one-to-one marketing. The model of this study will help e-commerce managers to decide whether a customer is good target or not in several days. In this study, data of an e-commerce site which sells golf products in Japan was used. However, customers will make two decisions, visiting and purchasing behaviour, at every e-commerce sites. Therefore, the model of this study will be applied to other e-commerce sites. As a marketing strategy, for "a customer who tends to purchase after viewing many product pages and doesn't tend to visit after viewing many product pages", prompting the customer to visit by sales promotion may be an example.

As a future study, the accuracy of prediction of purchase may need to be improved by the variables about the frequency of playing golf. Empirical analysis on other e-commerce sites such as a site which sells clothes will be needed and the result will be compared with the result of this study.

6. Acknowledgement

We would like to thank to professor Yamaguchi, T. (Department of Management Science, Faculty of Engineering, Tokyo University of Science) for his support on our experiment.

7. References

[1] Bucklin, R. E., Lattin, J. M. A Two-State Model of Purchase Incidence and Brand Choice. *Marketing Science.* 1991, 10(1): 24-39.

[2] Bucklin, R. E., Sismerio, C. Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach. *Journal of Marketing Research*. 2004, **41**(3): 306-323.

[3] Ministry of Economy, Trade and Industry of Japan, Retrieved September 27th , 2011 from the World Wide Web: http://www.meti.go.jp/policy/it_policy/statistics/outlook/h22houkoku.pdf

[4] Moe, W. ,Fader, S.P. Dynamic Conversion Behavior at E-Commerce Sites. *Management Science* . 2004, **50**(3): 326-335.

[5] Rossi, P.E., Allenby, G. and McCulloch, T. Bayesian Statistics and Marketing. John Wiley & Sons, Jersey. 2005.

[6] Van den Poel, D., Buckinx, W. Predicting online-purchasing behavior. *European Journal of Operational Research*. 2005,**166**(2): 557-575.