

A Study on Web Usage Behavior of Internet Shopping Mall User: W Cosmetic Mall Case

Hee Seok Song^a, and Hyung Chul Jun^b

^a*Department of Management Information Systems, Hannam University,
133 Ojung-dong, Taeduk-gu, Taejon, 306-791, South Korea
Tel: +82-42-629-7614, Fax: +82-42-629-7667, E-mail Address: hssong@hannam.ac.kr*

^b*Overseas Sales & Marketing System Team, AutoEverSystems
113-25, Wonhyo-ro, 4-ga, Yongsan-gu, Seoul, Korea
Tel: +82-2-3271-2801, E-mail: apple525@hanmail.net*

Abstract

With the rapid growth of e-commerce, marketers are able to observe not only purchasing behavior on what and when customers purchased, but also the individual Web usage behavior that affect purchasing. The richness of this information has the potential to provide marketers with an in-depth understanding of customer. Using commonly available Web log data, this paper examines Web usage behaviors at the individual level. By decomposing the buying process into a pattern of visits and purchase conversion at each visit, we can better understand the relationship between Web usage behavior and purchase decision. This allows us to more accurately forecast a shopper's future purchase decision at the site and hence determine the value of individual customers to the site. According to our research, not only information seeking behavior but also visiting duration of a customer and participative behavior such as participation in event should be considered as important predictors of purchase decision of customer in a cosmetic internet shopping mall.

Keywords:

Web Usage Mining, Web Usage Behavior, Purchase Decision, Participative Behavior

1. Introduction

In the context of Web usage mining research, Web usage behavior has played important role in building successful online retail strategies for Web site design, online advertising, market segmentation, product recommendation, and defection detection (Lohse et al. 2000; Datta et al. 2000; Mozer et al. 2000; Ng and Liu 2000; Raghavan et al. 2000; Song et al. 2001). In particular, Lohse et al. (2000) showed that the behavior of online customers is the most reliable information for future purchase decision rather than customers' stated intentions or the guesses offered by Web marketing experts. Despite the importance of Web usage behavior, we can not find any sophisticated model, which investigate relationship between Web usage behavior and purchase decision at the individual level. In other words, there have been limited studies that investigate on which

feature of usage behavior is important to customer purchase decision. Does the customer who stay longer ensure high likelihood of purchase? Does the customer refer Bulletin Board when they buy? This study aims to test on which behavior significantly affect individual purchase decision among various types of Web usage behaviors. For this purpose, we categorized Web usage behavior into visiting behavior (navigation and information seeking behavior) and participative behavior (Bulletin Board and event participation). Then, we tested relationships between those behaviors and purchase decision empirically for a women cosmetic mall. We evaluated our model in a specific shopping site rather than more general nature because of varieties in their structure and depth and width of information page.

2. Research Background

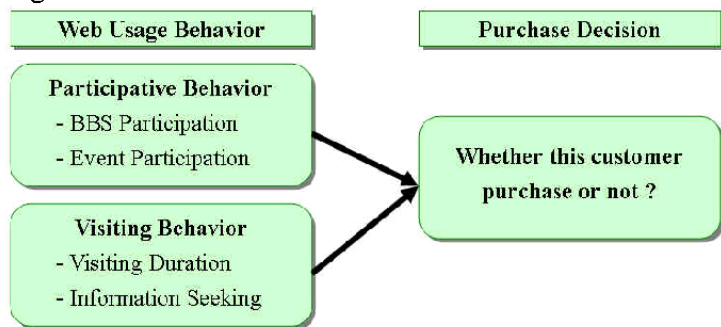
This study is based on the research of Web usage mining. Web usage mining is the application of data mining techniques to discover usage patterns from Web data, in order to understand and better serve the needs of Web based applications (Srivastava et al. 2000). The approach for Web usage mining mainly uses the log data directly by utilizing special pre-processing techniques. Log files are generated by certain software packages to record events occurred in a computer system. Because of the objectivity in its originating process, log data has become one of the most preferable sources that show how customers behave online. The technique has been used in studies of online information retrieval (IR) for a couple of decades, but it gained much popularity in 1990s with the widespread use of the World Wide Web. Although our study is based on the research of Web usage mining in its approach, it is different from the work of Web usage mining because this study aims to investigate the relationship between Web usage behavior and purchase decision not to develop user model or techniques.

3. Research Model

To investigate relationships between Web usage behaviors and purchase decision, we defined two categories of Web usage behavior which can be extracted from log files: participative behavior and visiting behavior.

Participative behavior is a proactive behavior which is distinguished from simple visiting behavior. Participative behavior may be positively related to customer purchase decision. Koh and Kim (2003) provided significant empirical results from virtual community that knowledge sharing activity is positively related to community promotion and customer loyalty. And they expected that such loyalty results ultimate commercial performance. In our site, BBS(Bulletin Board System) and event participation are adopted to measure participative behavior. These two construct are commonly available information in most online shopping mall. BBS participation mainly includes knowledge posting and viewing activities. As a measurement of BBS participation, we extracted information on whether he or she referred BBS for the purpose of reading or writing during unique visit. For the event participation, we extracted information on whether he or she participated in company driven event during his visit.

We chose visiting duration and information seeking behavior as a construct for simple visiting behavior. Information seeking behavior is expressed by clicks of product pages. Heavy information seeking behavior may be positively related to purchase decision because many studies apply number of clicks about product pages as an index of product preference or buying intention (Cooley et al. 1999b; Langley 1999). Also, the longer customer stays, the higher probability of purchase is expected. Kim et al. (2001) pointed that page view has been focused frequently to find out which page draws more attention in the previous Web usage mining but duration in this page hasn't drawn attention. They emphasized the importance of studying spent time on each page. We extracted session time and the number of clicked pages during a visit as a measurement of visiting duration and information seeking behavior. The definition of session time is explained in the next section. Figure 1 summarizes our research model.

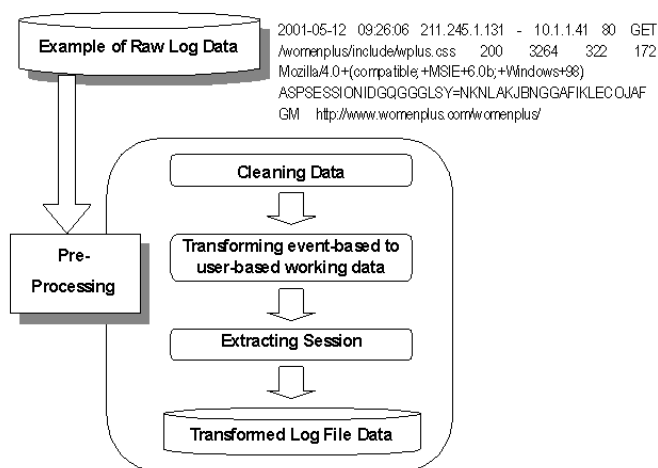


[Figure 1] Research model

4. Domain and data pre-processing

W cosmetic mall is providing information on beauty and selling cosmetics as well as supplying cyber place for communication for female Internet user since 1994. It ranked the fourth most frequently visited site among Korean female Web sites by a survey carried during 28th March through 16th April 2001. Around 3000 Internet users visited daily and 100 users among them purchased more than one product daily during examined dates. We explain the procedure for pre-processing of log because this study adopt

the approach of Web usage mining in data collection process. The pre-processing steps include the domain dependent tasks of data cleaning, user identification, session identification and path completion as explained in Figure 2 (Srivastava et al. 2000).



[Figure 2] Pre-processing procedure

For this step, we extracted Web log from three chosen days (12th and 25th May and 7th June, 2000) and identified user and session through above mentioned procedure. After this screening, we had 613 sessions whose average logical session times is 30.98 minutes, which lies in the average Internet access time zone of Korean users. Table 1 shows descriptive statistics on customers' session time. 313 sessions are 'ordered' session and 300 sessions are 'non-ordered' session. These 300 were selected for comparison sample set, by stratified random sampling. Because 'Non-order' sessions vary from 1 to 5000 seconds in their length, we stratified them into 50 strata and picked out samples in accordance with the density of each stratum

[Table 1] Descriptive Statistics for Session time

Descriptive Statistics (seconds)	
Mean	1858.590538
Standard Error	69.4998109
Median	1486
Mode	2189
Standard Deviation	1720.734476
Pooled Variance	2960927.138
Minimum	6
Maximum	10238
Count	613

5. Hypothesis Analysis

We analyzed relationship between participative behavior and purchase decision. We tried to verify whether a customer who is eager to participate in site event or BBS activity actually orders more or not. Thus we hypothesise as follows.

H1-1: A customer refers BBS when he or she buys

The chi-square test is appropriate for situations in which a test for differences between samples is required (Cooper et al. 1995). We checked both 'ordered' and 'non-ordered' customer groups whether they had referred or participated in BBS. The results appear in the following Table 2.

[Table 2] Contingency Table of Chi-Square Result on BBS-Order Relationship

Count Expected Values	BBS Referred	BBS Non Referred	Row Total
Ordered	159 149.27	141 150.73	300
Non-Ordered	146 155.73	167 157.26	313
Column Total	305	308	613

- Hypothesis.**
 H_0 : There is no difference in order issue between 'BBS referrers' and 'non-referrers.'
 H_1 : There is a difference in order issue between 'BBS referrers' and 'non-referrers.'
- Statistical Test.** χ^2 is appropriate.
- Significance level.** $\alpha = .05$ with d.f.=(2-1)(2-1)=1.
- Calculated value.** $\chi^2 = 2.4743$
- Decision.** $p=0.1157$ ($\alpha = .05$, d.f.=1) and there's no reason to reject null hypothesis.

It means there is no typical relationship between 'BBS referrers' and 'non-referrers' in order issue. We tried additional test because this was an unexpected result. A customer spends more time buying a luxury vehicle than a paper towel. Price drives various degrees of risk when a purchase would be fail. Valuable products naturally ask a customer to consider deeply and to search more information on the item than cheap ones (Newman and Staelin 1972). In a Web site, it is possible to assume a customer refers more in BBS when he or she purchase a high price product. Therefore we tested modified hypothesis as follows.

H1-1': A customer refers BBS when he or she buys valuable product

Customers are highly involved with an expensive, infrequent or risky purchase (Kotler and Armstrong 1999). We found prices for each product in integrated log file and categorized them into 4 different price groups and performed chi-square test using contingency table again. As a result, we can't reject null hypothesis also. It means there is no typical relationship between 'BBS referrers' and 'non-referrers' although customer buy valuable product. The second hypothesis is relationship between event participation and purchase decision

H1-2: A customer participates in event when he or she buys

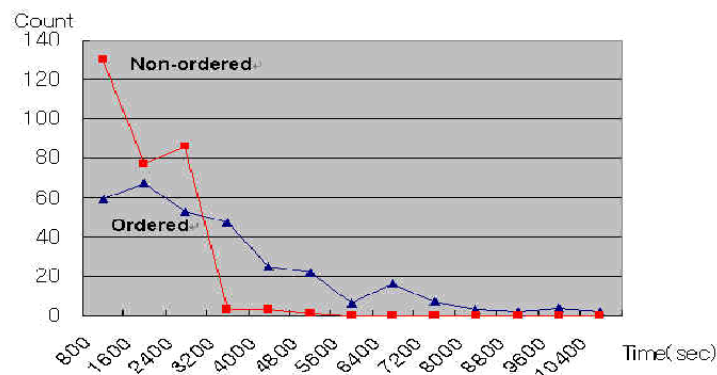
Event participation can be also examined with chi-square test. From the result, we can reject null and there is typical relationship between event participation and order issue. This is because most events appear in splendid pop-up window to attract customers' attention and provide

considerable price reduction.

It is commonly believed that the increase in session time length helps sales in online store. Web site owners make efforts to increase the stickiness of their sites. However, it is an unproven hypothesis that stickiness actually increases the sales in online store (Lee et al. 2001). As stated above, there have been several works showing that the longer a customer stays in a Web site the higher the probability he or she purchase a product in that site. However, few trials have been performed focusing on this hypothesis especially using statistical method.

H2-1: The longer a customer session becomes, the higher the purchase probability

Figure 3 shows that most 'non-ordered' customers leave the site before 2600 seconds or about 40 minutes while 'ordered' customers stay until 7200 seconds or 120 minutes roughly. And it also shows that session time can be one of the specific measurement when we observe user behavior related with purchase in a Web site.



[Figure 3] Comparison of session time

Two experiments were conducted to determine if the hypothesis just discussed is valid or not. Because we're not sure of variances of two samples, we've conducted t test under the assumptions of unequal variance. Decision is to reject H_0 and conclude that the mean session time of customers who ordered is longer than that of who didn't order. The result means that the mean session time of customers who ordered is longer than that of who didn't order under unequal variances assumption. As proved above, length of session time has strong power to tell if a customer would buy a product or more in a Web site or not. Although length of session time is rather user specific, it is clear that a site manager has to detain a customer as long as possible to drive the customer to purchase from the result of H2-1. The final hypothesis is relationship between information seeking behavior and purchase decision

H2-2: The more a customer refers, the higher the possibility of purchase becomes

As mentioned previously, current Web site evaluation sites are using page views as their yardstick. We counted only the pages which contain product information excepting the pages in order process. It is natural that 'ordered' customers need to refer more pages than 'non-ordered' when we count

ordering pages also. And we exempted non product pages because most of them are clicked for the purpose of navigation. Because the independent variable is nominal, it is possible to use logit test to check validity of H2-2. The estimated logit response model is as follows.

$$\text{Logit}(P_i) = -0.2264 + 0.00753X_1$$

With 95% confidence level, we reject the null hypothesis, every beta is equal to zero, because the calculated t is greater than any t value for 1 degree of freedom and $\alpha = .05$.

However, with 99% confidence level, we can't reject null. This means that clicked page can't explain the customers' order with 99% confidence level. This results show that page view alone can't tell a customer's purchasing behavior in high confidence level. This is an important result for site managers because the most site ranking is performed only based on page view.

6. Conclusion

Information about how customers buy is critical to build marketing strategy. Sales forces have long been focused on behavioral segmentation and it's still valuable way to distinguish different types of customers (Lochridge 2001). For the purpose of identifying specific Web usage characteristics of customers before they buy in a Web site, we evaluated the relationship between Web usage behavior and purchase decision using log files from a Korean cosmetics Web site. And we categorized Web usage behavior into visiting behavior during navigation and information seeking and participative behavior such as BBS(Bulletin Board System) and event participation. From the test result of hypothesis, we found visiting duration (or session time) and event-participation as a nice identifier that delimits 'non-ordered' customers from 'ordered'. We found that page click would be fail to distinguish 'ordered' customer from 'non-ordered' customer with 99% of confidence level in a cosmetic Web shopping mall. Therefore, we conclude that duration or customer session time, number of clicked pages and event-participation should be considered as the predictors of sales performance for this site. As a further research, we have a plan to overcome our limitation. It is required to test our model under more general nature. It is necessary to control the affection of extraneous variables which are occurred by differences among Web sites.

References

Cooley, R., Mobasher, B., and Srivastava, J. (1999a), "Data preparation for mining World Wide Web browsing patterns", *Knowledge & Information Systems* 1(1), pp. 5-32.

Cooley, R., Mobasher, B., and Srivastava, J. (1999b), "Discovery of Interesting Usage Patterns from Web Data", *Proceeding of WEBKDD '99*.

Cooper, Donald R., and Emory, C. W. (1995), *Business Research Methods*, 5th ed., IRWIN.

Datta, P., Masand, B., Mani, D. R., and Li, B. (2000), "Automated Cellular Modeling and Prediction on a Large Scale", *Artificial Intelligence Review* 14, pp. 485-502.

Kim, J. H., and Kim, J. R. (2001), "Analysis on Web user access pattern using association rules based on time", *Proceeding of Joint Workshop of Korea Management Science and Industrial Engineering Society*, pp. 852-855.

Koh, J., and Kim, Y. G. (2003), "Knowledge sharing in virtual communities: an e-business perspective", *Expert systems with applications* 26(2), pp. 155-166.

Kotler, P., and Armstrong (1999), *Principles of Marketing*, 8th ed., Prentice-Hall.

Langley, P. (1999), "User modeling in adaptive interfaces", *Proceeding of the seventh international conference on user modeling*, pp. 357-370.

Lee, J., and Podlaseck, M., Schonberg, E., and Hoch, R. (2001), "Visualization and Analysis of Clickstream Data of Online Stores for Understanding Web Merchandising", *Data Mining and Knowledge Discovery* 5 (1-2), pp. 59-84.

Lohse, G. L., Bellman S., and Johnson, E. J. (2000), "Consumer Buying Behavior on the Internet: Findings from Panel Data", *Journal of Interactive Marketing* 14(1), pp. 15-29.

Masand, B., and Spiliopoulou, M. (2000), "WEBKDD-99: Workshop on web usage analysis and user profiling", *SIGKDD Explorations* 1(2).

Moe, W. W. (2000), "Which visits lead to purchases?", Ph. D. Dissertation, University of Pennsylvania.

Mozer, M. C., Wolniewicz, R., Grimes, D. B., Johnson, E., and Kaushansky, H. (2000), "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry", *IEEE Transactions on neural networks* 11(3), pp. 690-696.

Newman, J. W. and Staelin, R. (1972), "Prepurchase information seeking for new car and major household appliances", *Journal of Marketing Research* 13, pp. 237-244.

Ng, K., and Liu, H. (2000), "Customer retention via data mining", *Artificial Intelligence Review* 14(6), pp. 569-590.

Raghavan, N., Bell, R. M., and Schonlau, M. (2000), "Defection Detection : Using online activity profiles to predict ISP customer vulnerability", *Proceedings of the sixth International Conference of Knowledge Discovery and Data Mining*, pp. 506-515.

Song, H. S., Kim, J. K., and Kim, S. H. (2001), "Mining the change of customer behaviour in an Internet shopping mall", *Expert systems with applications* 21(3), pp. 157-168.

Spiliopoulou, M., Pohle C., and Faulstich L. C. (1999), "Improving the Effectiveness of a Web Site with Web Usage Mining", *Proceeding of WEBKDD '99*.

Srivastava, J., Cooley, R., Deshpande, M., and Tan, P. N. (2000), "Web usage mining: Discovery and applications of usage patterns from web data", *SIGKDD Explorations* 1(2).

Yu, L., and Apps, A. (2000), "Studying e-Journal User Behavior Using Log Files: The Experience of SuperJournal", *Library & Information Science Research* 22(3), pp. 311-318.