**What Are They Trying to
Accomplish on my Website, Now?
Dynamic Identification of
Anonymous Consumers' Visit Goals**

S. Senécal, P.J. Kalczynski,
M. Fredette

What Are They Trying to Accomplish on my Website, Now?

Dynamic Identification of Anonymous Consumers' Visit Goals

Sylvain Senécal

Department of Marketing
HEC Montréal
sylvain.senecal@hec.ca

Pawel J. Kalczynski

Department of Information Systems and Decision Sciences
California State University, Fullerton
pkalczynski@fullerton.edu

Marc Fredette

GERAD and Department of Management Sciences
HEC Montréal
marc.fredette@hec.ca

October 2008

Abstract

This paper presents a model for identifying general goals of anonymous consumers visiting a retail website. When visiting a transactional website, consumers have various goals such as browsing or purchasing a particular product during their current visit. By predicting these goals early in the visit, online merchants could personalize their offer to better fulfill the needs of consumers. Most visitors remain anonymous to the website, however personalization systems require demographic and transaction history data which is available only for registered and logged-on users. We propose a simple model which enables classifying anonymous visitors according to their goals after only a few traversals (clicks). The model is based solely on navigational patterns which can be automatically extracted from clickstream. Theoretical and managerial implications are presented.

Key Words: Clickstream, consumer behavior, e-commerce, goal, anonymous visitors, personalization, logistic regression.

Résumé

Cet article présente un modèle pour l'identification de l'objectif d'un consommateur visitant un site Web. Lors de la visite d'un site Web transactionnel, les consommateurs peuvent avoir divers objectifs tels que la navigation ou l'achat d'un produit en particulier. En prédisant l'objectif d'un consommateur au début de sa visite, les commerçants en ligne peuvent personnaliser leur offre pour mieux répondre aux besoins de ce consommateur. Nous proposons un modèle simple qui permet de classer les visiteurs en fonction de leur objectif après seulement quelques clics. Le modèle est fondé uniquement sur des types de navigation qui peuvent être extraits automatiquement à partir des clics de souris. Les implications managériales et théoriques sont présentées.

1   Introduction

Consumers may have different goals while visiting a website (Moe, 2003). For instance, a consumer can visit a retailer's website to identify the most important attributes to look for when shopping for a product (e.g., megapixels for a digital camera). Another visitor could enter the same website to compare two competing products (e.g., Sony X vs. Pentax Y), while someone else could visit the website to buy a specific product (e.g., Nikon Z). Hence, the website would have hosted three visitors with three different visit goals. On most websites these three visitors would interact with the same content. This may be one explanation to why so many consumers are dissatisfied with transactional websites, e.g., (DoubleClick, 2004).

Personalization is one avenue that promises to improve customer satisfaction with websites. Online personalization gives retailers two major benefits. First, it allows them to provide accurate and timely information to customers which, in turn, often generates additional sales (Postma and Brokke, 2002). Second, personalization has also been shown to increase the level of loyalty that consumers hold toward a retailer (Srinivasan et al., 2002). Furthermore, Wind and Rangaswamy (2001) suggest that among all possible advantages of electronic commerce for retailers, the capacity to offer consumers an adaptive and personalized relationship is probably one of the most important ones.

Even though personalization systems are generally considered successful, they have serious limitations. Most of these systems use collaborative filtering techniques which require their (registered) users to log on in order for the system to access their demographic data and transaction history. Collaborative systems usually identify neighbors, i.e., people with potentially similar interests, for a given customer and use nearest neighbors' choices to provide recommendations. Finding a neighborhood for a given customer requires additional information such as demographics, purchase history, for a large number of users (including the customer in question) and a large number of non-binary product ratings (Cho et al., 2002). In addition, instead of indicating current needs or interests of a visitor entering a website, such recommendations usually represent interests of registered neighbors of a logged-on customer averaged over some period in time.

While the ratio of anonymous (unauthenticated) website visitors versus all visitors varies across websites and depends on many factors, it is safe to assume that the vast majority of retail website visitors remain anonymous until a transaction is made. In addition, most registered customers log on only after a transaction is about to be performed so they remain anonymous early in the site session (unless a tracking cookie is present in their system). Serving these visitors, which are often referred to as "non-transactional visitors," remains a challenge because of the lack of information available for authenticated users (Albert et al., 2004). Furthermore, tracking non-transactional visitors using cookies may be unreliable because it does not distinguish individuals using the same browser (account) and it does not recognize the variety of needs of individual visitors across visits (e.g. buying gifts vs. buying for oneself) and over time (Shahabi and Banaei-Kashani, 2003).

Mobasher et al. (2002) observed that anonymous clickstream data can provide enough information about the site session for making recommendations early in this session; they indicated that such solutions can help retain and convert unauthenticated visitors. If a retailer could identify the different goals of its anonymous visitors after only a few clicks, it could personalize its offers and, thus, consumers would be better served which, in turn, would improve their satisfaction and potentially the website's conversion rate. Albert et al. (2004) and Shahabi and Banaei-Kashani (2003) suggest that personalization strategies for anonymous visitors should be designed in an online rather than offline environment.

The objective of this research is to dynamically identify website visit goals of anonymous consumers while they are navigating through the website. More specifically, our objective is

to use clickstream data to categorize (and re-categorize) website visitors as either having a purchasing goal (i.e., shoppers) or a non-purchasing goal (i.e., browsers) after each traversal (click) they make on the website, without any additional information about the visitors (e.g., demographic data, purchase history, product ratings, or information stored in cookies).

In the next section, a review of the relevant literature is presented, and then research hypotheses are posited. Next, the methodology section explains how the clickstream data was collected, transformed, and used to test our hypotheses and model. After that, the results of the analysis are presented. Finally, the discussion section at the end of this paper highlights the main findings, contributions to research and practice, limitations, and future research avenues.

2 Conceptual Foundations

2.1 Consumer Goals

In this paper we investigate visitors' goals rather than outcomes of online sessions. The relationship between goals and behaviors has a long research tradition in psychology; see e.g. (Austin and Vancouver, 1996) for a review. For instance, Miller et al. (1960) suggested a framework (i.e., the TOTE cycle), in which goals play a significant role in explaining human behaviors. Although scholars recognize the utmost importance of goals in consumer research, the research on consumer goals in marketing is still sparse; see (Dholakia and Bagozzi, 1999) for a review.

Goals are "internal representations of desired states, where states are broadly construed as outcomes, events, or processes" (Austin and Vancouver, 1996). While visiting a website, consumers may have different goals such as seeking product or company information, purchasing a product, or posting a testimonial. Different goals have been shown to lead to different behaviors. For instance, consumers with different shopping goals acquire different pieces of information (Huffman and Houston, 1993). In their study, Huffman and Houston (1993) show that consumers with a specific goal (e.g., seeking product comfort) acquire more information on attributes that are relevant to their goal than on attributes related to other goals (e.g., seeking product versatility). Thus, the presence of a goal helps consumers structure their choice task. Furthermore, the degree of goal concreteness also has an effect on consumer information acquisition patterns. Peterman (1997) suggests that more concrete shopping goals (e.g., purchase a bicycle with excellent tires) lead consumers to acquire information across brands and that more abstract shopping goals (e.g., purchase a bicycle that is appropriate for commuting to work) lead them to seek within-brand information. The latter study also suggests that consumers with concrete shopping goals spend less time and acquire fewer pieces of information than consumers with more abstract shopping goals. In addition, consumers with concrete goals seem to spend more time processing each piece of information than consumers with abstract goals (Peterman, 1997). Furthermore, Lee and Ariely (2006) suggest that consumers have less concrete shopping goals at the beginning of their decision-making process and have more concrete goals and preferences as their shopping process progresses. They indicate that consumers' sensitivity to external cues (e.g., coupons or personalized offers) is likely to be higher at the beginning of the decision-making process since their goals are less well-defined at that moment (Lee and Ariely, 2006). Hence, it seems that personalization efforts should be made as early as possible when interacting with a consumer.

Consequently, anonymous consumers may have various goals when visiting an online retail store and these goals lead to different behaviors. Moreover, a consumer's goal could also evolve over the course of her/his shopping process. Thus, it becomes important not only to identify

anonymous consumers' website visit goals early in the session, but also to dynamically track them to identify any changes in these goals.

2.2 Consumers' Website Visit Goals

The classical consumer decision-making process suggests that consumers go through a series of steps (i.e., problem recognition, information search, evaluation of alternatives, intention, purchase, and post-purchase) while making a consumption decision (Engel et al., 1973). Nowadays, many consumers perform one or many of these decision-making steps online (Ratchford et al., 2003; Ratchford et al., 2007). As mentioned in the opening vignette, consumers with very different goals (from the information search to post-purchase activities) visit the same website. Hence, an online retailer has to find ways of fulfilling its visitors' needs in order to help them achieve their visit goals. One type of data that is readily available to online retailers is the Web-usage log. Using this data, a retailer can personalize its website by analyzing the clickstream of each visitor in real time.

Marketing researchers are increasingly using clickstream data, which was originally collected for website performance analyses. Clickstream has been used to investigate consumer behaviors across websites (Goldfarb, 2002; Johnson et al., 2003; Johnson et al., 2004; Park and Fader, 2004) and within specific websites (Kalczynski et al., 2006; Montgomery et al., 2004; Sismeiro and Bucklin, 2004). In the latter category, some studies focused on single visits to a given website, e.g., (Kalczynski et al., 2006), some dealt with multiple visits, e.g., (Bucklin and Sismeiro, 2003), while others investigated visits of both types, e.g., (Moe and Fader, 2004). Researchers engaged in this type of work have focused on such issues as: (1) identifying which visitors are likely to make a purchase (Moe and Fader, 2004), (2) information search and usage (Johnson et al., 2004), (3) the question of why consumers continue browsing on a website (Bucklin and Sismeiro, 2003), and (4) online decision-making processes (Senecal et al., 2005).

It has been suggested that visitors of online retail websites can be classified into four different categories based on clickstream data (Moe, 2003). Two of these categories encompass browsers (i.e., visitors with no intention to buy, such as knowledge builders and hedonic browsers) and the remaining two categories represent shoppers (i.e., visitors with immediate or future buying intentions, such as directed buyers and searchers/deliberators). Important clickstream-related differences were found between browsers and shoppers. As expected, shoppers show higher conversion rates than browsers, but they also seem to consult a different set of pages, revisit more product pages, and use the website search engine more often than browsers (Moe, 2003). In addition, based on their analysis of clickstream data, Montgomery et al. (2004) suggest that, during a visit to a retailer's website, consumers can switch from a browsing state to a deliberation state, thus evolving from a browser to a shopper within the same visit. Similarly to Lee and Ariely (2006), Montgomery et al. (2004) observe that consumers can start their website visit with one goal in mind and modify their goal along the way due to some external cues on the website. For their part, Kalczynski et al. (2006) use clickstream data to categorize visitors based on the probability of completing an online purchase. Using clickstream data from various websites they classify visitors as either buyers or non-buyers with only two variables: the linearity and density of visitors' navigational patterns.

The above research suggests that consumers' decision-making process is shaped in part by the goals pursued by the consumers. The Web-usage log is based on the actual behavior of website visitors and, as such, it is a promising source of data that can be transformed and used to assess consumers' goals. What is more, the dynamic analysis of clickstream data enables re-assessing the visit goal after each click performed by the anonymous consumer; it may help track potential changes to consumers' visit goals.

3   Hypotheses

Peterman's (1997) results seem to indicate that consumers with a clearer goal spend more time per piece of information. In our context, this finding suggests that shoppers (more concrete visit goal) should spend more time per page than browsers (more abstract visit goal). Furthermore, Moe (2003) suggests that online shoppers revisit more pages than browsers. Thus, the following hypotheses are posited.

Hypothesis 1: While visiting a retail website, shoppers spend more time per page on new pages than browsers.

Hypothesis 2: While visiting a retail website, shoppers spend more time per page on revisited pages than browsers.

Following Huffman and Houston (1993) and Moe (2003), we suggest that since shoppers have different visit goals, they will also have a different information search strategy than browsers and that these differences will show in their usage of the website search engine. Similarly to Moe (2003), we suggest that, on average, shoppers should use the website search engine more frequently. The following is proposed.

Hypothesis 3: While visiting a retail website, shoppers use the website search engine more frequently than browsers.

Based on these hypotheses, a model is proposed in order to discriminate between visitors with a purchasing goal (i.e., shoppers) and visitors with no purchasing goal (i.e., browsers). Hence, three variables will be introduced in the model: the time spent on new pages (i.e., Forward navigation, "F"), the time spent on already visited pages (i.e., Backward navigation, "B"), and the number of searches using the website search engine ("S").

The proposed model distinguishes itself from previous approaches in the following two ways: (1) it provides results early in the session, thus enabling taking appropriate actions before the visitor abandons the website and (2) no additional knowledge about the consumer, other than the navigational pattern automatically extracted from clickstream, is assumed. These two properties enable practical applications of the proposed model to sessions which can not be handled by traditional recommender agents. This particularly refers to cases, in which the identification of the visitor is not possible (e.g., the visitor did not log on) or the identified customer exhibits atypical behavior (e.g., he or she is shopping for someone else). In addition, the proposed model contributes to the theory of e-commerce by demonstrating how much information can be extracted from the content-independent Web-usage log without employing sophisticated and resource-consuming systems.

4   Methodology

4.1   Sample and Procedure

One hundred and thirty eight consumers were recruited from a consumer panel. All participants first received an email invitation and then were re-contacted to schedule a meeting at their convenience. The user data was collected in a laboratory setting at a major North American University. Once arrived at the laboratory, consumers were told that they would have to answer a questionnaire, navigate through a local online music store, and then complete another questionnaire. All participants were given an electronic gift certificate redeemable on this website in exchange for their participation. All subjects were assigned to either a shopping task or a browsing task. In the browsing task (TASK A), the subjects were asked to

browse through the website and be prepared to answer a few questions about its content; they were given an electronic gift certificate after they completed the final questionnaire. For the shopping task (TASK B), the subjects were asked to shop for a music CD of their choice and buy it with the electronic gift certificate if they found what they wanted. If not, they were told to keep the gift certificate for a future purchase. In order to maximize the involvement of the subjects in the shopping task, before going online they were asked which music genre and artist they would shop for. Seventy three percent reported a music genre that they would be interested in and sixty percent reported a specific artist.

Contrary to previous studies using clickstream data to investigate consumers' goals or states (Moe, 2003; Montgomery et al., 2004), we do not infer visit goals from clickstream data. Instead, we manipulated participants' goal, which insures that the observed clickstream data is precisely related to a specific visit goal.

Participants were relatively young (84% were less than 35 years old) and the sample was composed of students (60%) and workers (40%). Seventy two percent of participants had more than five years of experience with using the Internet. In order to rule out any differences between the two groups of participants (i.e., shoppers and browsers) caused by external factors, participants were asked to answer questions about their knowledge of their favorite music genre, the Internet, and their attitude toward the retailer. Participants' subjective knowledge of the Internet and of their favorite music genre was assessed using Flynn and Goldsmith's (1999) measurement scale (Cronbach alpha = 0.827 and 0.838 for the Internet and their favorite music genre, respectively) and their attitude toward the retailer was assessed using McKenzie and Lutz's (1989) measurement scale (Cronbach alpha = 0.943). No differences between shoppers and browsers were found relative to their subjective knowledge of the Internet (F= 0.129, p= 0.720), their subjective knowledge of their favorite music genre (F= 0.487, p= 0.487), and their attitude toward the retailer (F= 0.683, p= 0.410).

4.2 Clickstream Measures

Two time stamps (page displayed, traversal initiated) and the URL of every page visited were collected for each individual session. The final clickstream dataset consisted of 138 sessions (2,798 traversals). Out of the 96 participants assigned to the shopping task, 13 completed a purchase. The remaining shoppers decided to stop their session because they did not find the product they wanted. The average number of clicks was about 17 per session for TASK A (browsing) and about 22 for TASK B (shopping); the number of clicks per session ranged from 4 to 105. Figure 1 shows the relative frequency distributions of the number of traversals (clicks) per session for tasks *A* and *B* respectively. One can observe that most sessions consisted of 10 to 30 clicks.

4.3 Data Transformation

Before raw clickstream (Web-usage) data can be used in models, the unit of analysis has to be chosen and the data has to be transformed into variables. Transformation (aggregation) is necessary as clicks are not independent of each other (Zheng et al., 2003). Mobasher et al. (2002) emphasized that clickstream data has to be properly aggregated in order to be useful (i.e., actionable); they called the aggregates "aggregate usage profiles."

In this paper we use a technique known as "clipping at every click" (Van der Meer et al., 2000), that is, we analyze every *site session* (W3C, 1998) in an incremental manner, computing aggregate usage profiles after each consecutive click (traversal). Given the dynamic aspect of our model, clipping seems to be the most appropriate technique because inferences are made by the system after each traversal (click).
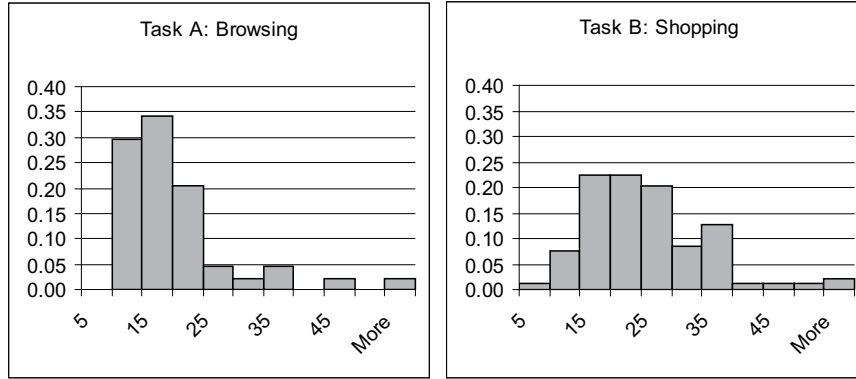
Figure 1: Relative Frequency Distributions of the Number of Traversals per Session

For the purpose of this research, each traversal was marked as forward (F), backward (B), or search (S). The forward traversal results when the visitor chooses a hyperlink leading to the new (previously unvisited) content, whereas a backward traversal indicates that the visitor revisited a page while going "backward" on the website. The search traversal results when the consumer chooses to use the search engine, thus bypassing the navigational structure. Figure 2 shows sample clickstream data tagged as F, B, or S. The numbers in parentheses indicate the number of times a given node (page) was visited. For instance, the first clickstream sequence indicates that the participant visited two new pages, used the search engine, then moved back to the previously-visited page, then used the search engine again, etc.

The proposed transformation resulted in a simple representation of the navigational paths taken by the visitors to accomplish their goals.

4.4 The Model

In order to classify visitors as either browsers or shoppers, we assumed that the differences in navigational patterns between these two groups can be measured using the number of times the search engine was used (H3) and the time spent viewing content pages after moving forward (H1) or backward (H2) on the website. If the goal remains constant, then one can expect that shoppers will use the search engine more often and spend more time reading the content of each webpage early in the session (Moe, 2003). On the other hand, browsers are likely to use the search engine less often and spend less time reading the content of the pages visited early in the session (Moe, 2003). Below we describe a model capable of classifying visitors as browsers or shoppers using clickstream data collected after k traversals.

Let m be the total number of sessions recorded. Let n_i denote the number of pages visited by the visitor in the i-th session (including the starting page). Let t_{ij} be the time spent by the visitor viewing the j-th page ($j = 1, \ldots, n_i$) accessed in the i-th session ($i = 1, \ldots, m$).

> *FFSB(2)SFSSSSFFSFB(2)FFFFFFFFFFF*
> *FFFFFB(2)FFFB(2)FFFB(3)FFFFFFFF*
> *FFFFFB(2)FB(2)B(2)SB(3)B(2)B(3)*
> *FFSFFFFSFSFB(2)FB(3)FB(4)FSSSFSSSS*

Figure 2: Sample Clickstream Data Tagged as Forward, Backward, or Search Traversals

For each individual session i, and each individual traversal $k < n_i$, one can compute the total time spent on the website after the k-th traversal: $T_i^k = \sum_{j=1}^{k+1} t_{ij}$. Further, let TF_i^k denote the total page-viewing time after moving forward computed after the k-th traversal in the i-th session. Similarly, let TB_i^k be the total page-viewing time after moving backward computed after the k-th traversal in the i-th session. Also, let CS_i^k denote the number of times the search engine was used during the first k traversals in the i-th session. For example, $TB_1^5 = 67$ indicates that, after five traversals, the visitor in session 1 spent a total of 67 seconds viewing previously-visited pages.

We propose the following binary logistic regression model to classify visitors as shoppers or browsers after the k-th traversal:

$$\pi_i = 1 \Big/ \left(1 + e^{-\left(\beta_0 + \beta_1 TB_i^k + \beta_2 TF_i^k + \beta_3 CS_i^k\right)} \right), \tag{1}$$

where and β_0, β_1, β_2, and β_3 are the coefficients of the model, e is Euler's number, and π_i denotes the probability that the i-th session is a shopping session.

5 Results

Using binary logistic regression, we estimated the parameters of the model for different values of k (ranging from 2 to 10) using the available data. Smaller values of k indicate the "early in the session" period, which is the most interesting from a practical standpoint. Table 1 presents the summary of the results of fitting of the proposed model.

Each regression model yielded a non-significant Hosmer and Lemeshow's goodness of fit test (see Table 1), which suggests that there are no significant differences between the observed and predicted classifications (Hair et al., 2005).

The odds ratios presented in Table 1 are greater than one; this shows that participants with a shopping visit goal spent more time per page than participants assigned to the browsing

Table 1: Logistic Regression Results

Session clicks	2*	3*	4	5	6	7	8	9	10
Hosmer-Lemeshow	0.30	0.10	0.10	0.48	0.78	0.43	0.73	0.31	0.99
Nagelkerke R2	0.27	0.33	0.30	0.35	0.41	0.42	0.42	0.44	0.45
Odds ratio (p<0.05)									
Time on page after Forward Click (F)	8	1.04	1.051	1.071	1.069	1.048	1.041	1.033	1.034
Time on page after Backward Click (F)	1.02	1.02	1.019	1.018	1.018	1.014	1.012	1.013	1.015
Number of Searches (S)	8	8	10.31	8.105	7.536	7.119	6.256	6.876	7.102
Percentage of Shoppers									
Naïve Rate	68.1	68.1	68.1	67.9	67.9	67.9	67.9	71.7	71.7
Percentage of Correct Classification									
Overall	**72.5**	**75.4**	**73.9**	**75.2**	**79.6**	**75.4**	**75.6**	**77.3**	**77.5**
Improvement over Naïve Rate (%) (overall-naïve)/(100-naïve)	**13.8**	**22.9**	**18.2**	**22.7**	**36.4**	**23.4**	**24.0**	**19.8**	**20.5**

Note: * The model can be used for prediction although some parameters are probably infinite
(possible quasi-complete separation of the data points).

goal. The former spent more time per page going forward (F) and backward (B) on the website, which confirms Hypotheses 1 and 2, respectively. Furthermore, it turned out that shoppers used the website search engine (S) more frequently than browsers, which supports Hypothesis 3.

Overall, the performance of the proposed model is satisfactory. Most importantly, the model is quite effective in classifying anonymous visitors as shoppers or browsers even after a very limited number of clicks. For instance, after only three clicks the model correctly classifies 75.4% of all the subjects. Between 2 and 10 clicks, the percentage of correct classification is between 72.5% and 79.6%. Moreover, between 2 and 10 clicks, the average improvement of correct classification over the naïve rate is 22.4%.

6   Discussion

In order to dynamically personalize offers, an online retailer needs to identify the visitor's goal early using the available data. Once the goal is identified (shopping or browsing in the present case) it is then possible to communicate personalized and, thus, more relevant information to the consumer. This paper demonstrates that clickstream data can be used to effectively classify anonymous visitors according to their goals. The three variables included in the model (time per page while going forward or backward and the number of searches) were found to effectively discriminate between anonymous consumers with a purchasing visit goal and anonymous consumers with a non-purchasing goal. Furthermore, our results suggest that it is possible to discriminate between these two groups of consumers as early as after two traversals.

6.1   Theoretical Implications

In contrast to past research on inferring consumers' website visit goals using clickstream data (Moe, 2003), in this study we manipulated the goals and then used clickstream data to build a predictive model that classifies consumers according to their goals. Thus, in addition to introducing a novel approach to building goal-predictive models using clickstream data, this paper contributes to research on consumer goals by unequivocally showing that online retail visit goals have a significant impact on consumers' online behaviors. In addition, this paper presents a very simple and effective model that can dynamically identify the goals of anonymous visitors early in their visit, which contrasts with more complex clickstream-based predictive models, e.g., (Kalczynski et al., 2006; Montgomery et al., 2004). Thus, we believe that similar models could be used to investigate other online goals and behaviors.

6.2   Managerial Implications

The output of our model could serve as the input to a recommender agent thus enabling better recommendations or dynamic personalization of webpage content according to the goal of an anonymous visitor. For instance, Sismeiro and Bucklin (2004) suggest that some website content (e.g., page clutter) may hinder the completion of certain tasks but, paradoxically, it can improve the completion of other tasks. Hence, the dynamic identification of anonymous visitors' goals as early as possible and adaptation of the content should help improve online task completion. Moreover, by being able to identify goals early in the visit, online retailers should be more effective in communicating their offers since, as suggested by Lee and Ariely (2006), consumers are more sensitive to external cues early in the shopping process.

The relative simplicity of the proposed model is advantageous as compared to more sophisticated clickstream prediction models. The data necessary for the proposed model (session

ID, time stamps, node ID) is collected in real-time by both IIS and Apache Web servers. In addition, since the proposed model is fairly simple, even small e-businesses could use it to predict visit goals of their anonymous visitors.

6.3 Limitations and Research Avenues

Researchers point to scalability as one of the major problems with using Web-usage mining for making recommendations; see, e.g., (Albert et al., 2004; Cho et al., 2002; Shahabi and Banaei-Kashani, 2003; Zheng et al., 2003). The volumes of Web-usage data collected by popular sites along with the need to provide recommendations online (during the session) typically result in the separation of the offline component from the online recommender component. The offline component utilizes Web-usage data to build or fit the models while the online one uses the model to make dynamic recommendations. The implementation of our model would likely follow this strategy.

Also, because we are unable to confirm whether our model works for other websites, consumers, and types of products/services, additional research should be conducted before applying these findings to business practice. We expect that the way in which the content is presented affects the time spent by visitors on each individual webpage and the number of times the website search engine is used.

If the model proves applicable to most websites, future research should focus on understanding the detailed intention of anonymous visitors to the website, i.e., the product, service, or piece of information that the visitors are interested in. In order to maximize the effectiveness of clickstream prediction models, content-independent (e.g., time per page) and content-dependent data (e.g., the label of the hyperlink clicked) need to be combined, see e.g., (Kalczynski et al., 2006; Moe, 2003; Montgomery et al., 2004). Thus, an interesting research avenue would be to incorporate content-dependent data in our goal-prediction model to improve its predictive power.

References

Albert T. C. Goes P. B. Gupta A. GIST: A Model for Design and Management of Content and Interactivity of Customer-Centric Web Sites. MIS Quarterly 2004; 28 (2): 161–182.

Austin J. T. Vancouver J. B. Goal Constructs in Psychology: Structure, Process, and Content. Psychological Bulletin 1996; 120 (3): 338–375.

Bucklin R. E. Sismeiro C. A Model of Web Site Browsing Behavior Estimated on Clickstream Data. Journal of Marketing Research 2003; 40 (3): 249–267.

Cho Y. H. Kim J. K. Kim S. H. A Personalized Recommender System Based on Web Usage Mining and Decision Tree Induction. Expert Systems with Applications 2002; 23: 329–342.

Dholakia U. Bagozzi R. P. Goal Setting and Goal Striving in Consumer Behavior. Journal of Marketing 1999; 63 (4): 19–32.

DoubleClick. Online Retailers Have Less Time to Persuade in e-Marketer. Double Click 2004 (November).

Engel J. F. Kollat D. T. Blackwell R. D. Consumer Behavior. Holt: Rinehart & Winston, 1973.

Flynn L. R. Goldsmith R. E. A Short, Reliable Measure of Subjective Knowledge. Journal of Business Research 1999; 46 (1): 57–66.

Goldfarb A. Analyzing Website Choice Using Clickstream Data: Elsevier Science Ltd., 2002.

Hair J. F. Black B. Babin B. Anderson R. E. Tatham R. L. Multivariate Data Analysis (6 ed.). Upper Saddle River, NJ: Prentice Hall, 2005.

Huffman C. Houston M. J. Goal-oriented experiences and the development of knowledge. Journal of Consumer Research 1993; 20 (2): 190.

Johnson E. J. Bellman S. Lohse G. L. Cognitive Lock-In and the Power Law of Practice. Journal of Marketing 2003; 67 (2): 62.

Johnson E. J. Moe W. W. Fader P. S. Bellman S. Lohse J. On the Depth and Dynamics of World Wide Web Shopping Behavior. Management Science 2004; 50 (3): 299–308.

Kalczynski P. J. Senecal S. Nantel J. Predicting Online Task Completion with Clickstream Complexity Measures: A Graph-Based Approach. International Journal of Electronic Commerce (IJEC) 2006; 10 (3): 121–141.

Lee L. Ariely D. Shopping Goals, Goal Concreteness, and Conditional Promotions. Journal of Consumer Research 2006; 34 (4): 556–566.

McKenzie S. B. Lutz R. J. An Empirical Examination of the Structural Antecedents of Attitude Toward the Ad in an Advertising Pretesting Context. Journal of Marketing 1989; 53 (2): 48–65.

Miller G. A. Galanter E. Pribram K. H. Plans and the Structure of Behavior. New York: Holt, Rinehart & Winston, 1960.

Mobasher B. Dai H. Luo T. Nakagawa M. Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization. Data Mining and Knowledge Discovery 2002; 6: 61–82.

Moe W. W. Buying, Searching, or Browsing: Differentiating between Online Shoppers Using In-Store Navigational Clickstream. Journal of Consumer Psychology 2003; 13 (1&2): 29–40.

Moe W. W. Fader P. S. Capturing Evolving Visit Behavior in Clickstream Data. Journal of Interactive Marketing 2004; 18 (1): 5–19.

Montgomery A. Li S. Srinivasan K. Liechty J. C. Modeling Online Browsking and Path Analysis Using Clickstream Data. Marketing Science 2004; 23 (4): 579–595.

Park Y.-H. Fader P. S. Modeling Browsing Behavior at Multiple Websites. Marketing Science 2004; 23 (3): 280.

Peterman M. L. The effects of concrete and abstract consumer goals on information processing. Psychology and Marketing 1997; 14 (6): 561–583.

Postma O. J. Brokke M. Personalisation in practice: The proven effects of personalisation. Journal of Database Marketing 2002; 9 (2): 137–142.

Ratchford B. T. Lee M.-S. Talukdar D. The Impact of the Internet on Information Search for Automobiles. JMR, Journal of Marketing Research 2003; 40 (2): 193.

Ratchford B. T. Talukdar D. Lee M.-S. The Impact of the Internet on Consumers' Use of Information Sources for Automobiles: A Re-Inquiry. Journal of Consumer Research 2007; 34 (1): 111.

Senecal S. Kalczynski P. J. Nantel J. Consumers' Decision-Making Process and Their Online Shopping Behavior: A Clickstream Analysis. Journal of Business Research 2005; 58 (11): 1599–1608.

Shahabi C. Banaei-Kashani F. Efficient and Anonymous Web-Usage Mining for Web Personalization. INFORMS Journal on Computing 2003; 15 (2): 123–147.

Sismeiro C. Bucklin R. E. Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach. Journal of Marketing Research. 2004; XLI (August 2004): 306–323.

Srinivasan S. S. Anderson R. Ponnavolu K. Customer loyalty in E-commerce: An exploration of its antecedents and consequences. Journal of Retailing 2002; 78 (1): 41–50.

Van der Meer D. Dutta K. Datta A. Enabling Scalable Online Personalization on the Web, Proceedings of: Electronic Commerce (EC00). Minneapolis, MN: ACM, 2000.

W3C. Web Characterization Activity Terminology Sheet, http://www.w3.org/wca/terminology.html, accessed on July 9, 2008.

Wind J. Rangaswamy A. Customerization: The next revolution in mass communicationJournal of Interactive Marketing 2001; 15 (1): 13–32.

Zheng Z. Padmanabhan B. Kimbrough S. O. On the Existence and Significance of Data Processing Biases in Web-Usage Mining. INFORMS Journal on Computing 2003; 15 (2): 148–170.