AMCIS 2002 Proceedings

Americas Conference on Information Systems (AMCIS)

December 2002

AN E-COMMERCE DECISION SUPPORT SYSTEM DESIGN FOR WEB CUSTOMER RETENTION

David Olson
University of Nebraska

Sebastian Elbaum
University of Nebraska

Steven Goddard
University of Nebraska

Fred Choobineh
University of Nebraska

Follow this and additional works at: http://aisel.aisnet.org/amcis2002

AN E-COMMERCE DECISION SUPPORT SYSTEM DESIGN FOR WEB CUSTOMER RETENTION

David L. Olson
University of Nebraska
Dolson3@unl.edu

Sebastian Elbaum
University of Nebraska
Selbaum2@unl.edu

Steven Goddard
University of Nebraska
Sgoddard2@unl.edu

Fred Choobineh
University of Nebraska
Fchoobineh1@unl.edu

Abstract

An advantage of E-commerce is the reduced cost in conducting transactions. However, a significant disadvantage is that many sales are lost that would not have been had the transaction been conducted person-to-person. The reasons for their early departure vary from poor quality of service, inability to find the desired product, to non-competitive prices. In the proposed system, the broker detects customer behavior indicating imminent departure, which triggers an update of the client's web page being displayed by the browser to maximize the likelihood of retaining the customer.

The proposed system includes the idea of an E-commerce broker to improve customer retention and to increase the profitability of sales. The broker sits between the client and the web server, acting as an intermediate agent that monitors the customer's behavior on the client side and notifies the server when the customer takes some action that may precede early departure form the site. In conjunction with on-line customer activity monitoring, a customer value model will be built for each customer with the intent of identifying site products most likely to be wanted by that customer.

Introduction

An emerging issue in electronic commerce is customer retention. This paper presents a system design for a decision support system intended to improve customer retention for web retail businesses. It seeks to enhance customer retention through monitoring of customer web-site visits to detect customers likely to leave, and through conjoint/multiple criteria analysis to assist in product pricing.

The future of any class of software product depends upon keeping up with new opportunities (Kumar and van Hillgersberg, 2000). The next generation of electronic commerce will involve revolutionary changes in services provided. Electronic tools also can provide more intelligent support to businesses through customer profiling, and development of models to anticipate customer behavior and identification of their preferences. Understanding customer value has become critical to contemporary business success (van Everdingen et al., 2000; DeSarbo et al., 2001).

Currently, on-line transactions are usually conducted without a sales agent via shopping-cart software. Even if an agent is involved, it is often through electronic communication and the agent seldom physically meets with the customer—hence, the term E-commerce. An advantage of E-commerce is the reduced cost in conducting transactions. However, a significant disadvantage is that many sales are lost that would not have been had the transaction been conducted person-to-person. The reasons for their early departure vary from poor quality of service, inability to find the desired product, to non-competitive prices. In the proposed system, the broker detects customer behavior indicating imminent departure, which triggers an update of the client's web page being displayed by the browser to maximize the likelihood of retaining the customer.

This approach was inspired by person-to-person selling practices. In person-to-person transactions, a sales agent is able to detect when a customer is losing interest in the product, and takes action to close the sale before the customer leaves. These actions can include price adjustments, enhancement of product packages, or other incentives to retain the customer. The sales agent is able to observe non-verbal signals displayed by the customer before the customer has consciously decided to leave without making a purchase. Good sales agents act on these non-verbal signals and change tactics to retain the customer. In a nutshell, our work seeks to incorporate some of the advantageous attributes of person-to-person transactions in on-line transactions.

The proposed system includes the idea of an E-commerce broker to improve customer retention and to increase the profitability of sales. The broker sits between the client and the web server, acting as an intermediate agent that monitors the customer's behavior on the client side and notifies the server when the customer takes some action that may precede early departure form the site. In conjunction with on-line customer activity monitoring, a customer value model will be built for each customer with the intent of identifying site products most likely to be wanted by that customer. This system can also be used to apply dynamic pricing. The E-commerce broker model can detect customer actions indicating a lost sale, and dynamically select content to post to the web interface. Policies adopted are selected on the basis of customer preference analysis. This paper presents this system in the form of an expected return model linking data collection, data analysis leading to identification of optimal policy, and policy implementation. While the system is automated, it falls within the purview of a decision support system by applying data and models to deal with a business decision.

Section 2 of the paper gives an overview of approaches others have applied to the problem, including web data analysis and conjoint analysis. Section 3 presents our E-commerce customer retention model, describing the architectures for both the conjoint model and the E-commerce broker model. Section 4 presents conclusions and directions for future research.

Related Work

Spiliopoulou et al. (1999) considered three aspects of using web site usage data for marketing analysis: 1) data acquisition; 2) cost and quality measures; and 3) assessment of user/owner satisfaction. Data acquisition can be invasive, actively contacting the user through questionnaires or other methods. Data acquisition can be non-invasive through recording user behavior without their interaction. Web server logs are the primary means of obtaining this information, which is often stored in web data warehouses for deep analysis. Cooley et al. (1999) discuss preparation of web log data for analysis.

Once data on customer web behavior is gathered, it can be used to measure satisfaction of the web site owner in various ways. Assessing user satisfaction with web sites has been an active field of research. Berthon et al. (1996) and Dreze and Zufryden (1998) assessed contact and conversion efficiency from web data. Eighmey (1997) and Dreze and Zufryden (1997) addressed value creation offered by a site from the perspective of the user. Dreze and Zufryden used conjoint analysis as a tool to measure web site effectiveness in delivering promotional content. Web-based data collection was found to be unobtrusive, experimentally based, externally valid, and easily capable of obtaining large sample sizes.

Web Data Collection

The opportunity to have a one-to-one interaction with customers that the Internet provides has resulted in a large amount of research on increasing profitability of an Internet shopping mall or site. First, many sites have been found to benefit from considering the importance of web-content's psychological effect on customers. Personalized web-content targeting specific customer preferences has led to a second level of web-competition. Third, dynamic web-content and personal discount offers can also be used by site managers. However, this third level of sophistication requires additional resources, and high-traffic sites may end up losing customers due to slower response time.

Web measurement has grown in technological ability as well as in application importance. Products such as that of Clickstream Technologies (www.clickstream.com) provide means of measuring web activity, much as Nielsen's measures television set activity. The most important uses of this type of measurement have been advertising statistics, checking web site audience growth, monitoring behavior of those visiting web pages, and monitoring web site operational management. The two basic methods of web measurement are to ask what users are doing, or to monitor what web servers are doing. Asking users is a form of panel-based research. Monitoring servers is server log analysis. Event logging is the process of recording user-initiated activities (www.netconversions.com). This practice has been widely developed. Event logs capture how web site users are using the interface. Table 1 provides comparison of the type of data captured by each method.

Table 1. Web Monitoring Data Captured

Server Logs	Event Logs
Date/time data	Path analysis
Entry/exit pages	Page abandonment
Requested pages	Field completion times
Downloaded files	Drop-down box selections
Types of browsers	Scrolling
Level of server activity	Click density by link, non-link
IP address	Mouse movements
Number of page views	Browser controls
Referring sites	Error occurrence

Adapted from Kangas & Chlu, 2001 [6]

Thus, it is clear that there is a great deal of information that can be obtained through web monitoring. This information has been widely used to analyze the effectiveness and weaknesses of web sites. Such information has been used to analyze web site visitor interests as well (Mobasher et al., 2000). Here we are proposing such technology be used for both forms of web data gathering: to elicit customer preferences, and to monitor web visitor behavior, specifically enabling measurement of the web visitor as a consumer.

Conjoint Analysis

Conjoint analysis (Green et al, 2001; Gustafsson, et al., 2001) derives preferences by asking respondents to choose between competing options with controlled differences in attributes, simulating the way people make choices when faced with multiple alternatives (Fraenkel et al., 2001). Montgomery (2001) contended that techniques developed outside the internet environment can be effectively utilized for internet operations, specifically with clickstream technology to obtain input data, to include conjoint analysis. Conjoint analysis is the use of consumer preference information to identify buyer trade-offs among competing products and/or services. Conjoint analysis includes a number of models to measure respondent preference orderings across multiattributed stimuli. Conjoint analysis has been widely applied in marketing research, and has been identified as the most used marketing research method for analyzing consumer trade-offs. Once a model is developed, it can be used not only for analysis of product design, but also to simulate consumer-purchasing behavior.

Green, et al. (2001) give four major types of preference input procedures used for conjoint analysis (see Table 2). All essentially reflect a utility function, where *k* attributes are evaluated on some scale (usually something like 0 to 100), and each attribute *i* is weighted, often with each weight constrained to between 0 and 1 and the sum of the weights equal to 1.0. The value to the consumer for every possible combination *j* of attribute values would then be expressed by some value function:

$$\text{value}_j = \sum_{i=1}^{k} \text{score}_{ij} \times \text{weight}_i$$

Full value preference models also include interaction terms. Full profile techniques involve presenting each respondent with a complete set of prop cards. The respondent sorts the cards into ordered categories, and then rates each card on a 0 to 100 likelihood-of-purchase scale. Compositional techniques require each respondent to rate the desirability of each set of attribute levels on a 0 to 100 scale and then rate the attributes on an importance scale. Hybrid techniques ask each respondent to perform a self-explicated evaluation as well as evaluation of a subset of full-profile cards. A utility function is developed by regression over the information obtained from these two tasks. Adaptive conjoint analysis first asks each respondent to perform a self-explication task, and then evaluates a set of partial-profile descriptions in pairwise combinations. Adaptive conjoint analysis requires fewer responses on the part of subjects. The full-profile and hybrid procedures were given by Green, et al. (2001) as the most popular in current practice. Adaptive conjoint analysis has recently proven popular as well, supported by internet software (Green and Srinivasan, 1990; Wittink and Cattin, 1998). For the context of this study, however, the compositional approach is most appropriate due to data collection capabilities.

Table 2. Conjoint Analysis Methods

Method	Description
Full profile	Complete sorting of all combinations of attributes; assessment on 0-100 scale
Compositional	Rate attribute levels on 0-100 scale; rate importances
Hybrid	Self-explicated evaluation; evaluation of subset of full-profile cards
Adaptive	Self-explicated evaluation; pairwise evaluation of attribute levels

In this research, the compositional form of conjoint analysis is proposed. Data to support this form of research can be obtained through surveys delivered over the web upon initial client contact, and periodically readministered to look for changes in customer preference. Additionally, this preference model for each customer can be checked against purchase behavior. If the model correctly predicts purchases, that is evidence of model accuracy. If purchase behavior is not accurately predicted by the model, that could trigger readministering the conjoint data collection.

E-Commerce Customer Retention Model

The model consists of two complementary components. The conjoint model based on survey input (which can be collected at the web site) develops an initial baseline customer preference model. The E-commerce broker model monitors web customer site clickstream behavior, with the potential of dynamically inserting web site content to improve the probability of customer retention. The conjoint survey input is straightforward. The conjoint model queries customers for relative attribute importance on initial visits, and subsequent visits triggered by variance between model forecasts and results. The conjoint model identifies available product scores from internal databases. This model can be used for pricing decisions, as well as a basis for policy identification for the E-Commerce Broker Model.

There is a training period to the E-commerce broker model, followed by implementation. In the training period, the e-broker system collects data from visitors to the target web site. This data is analyzed to enable identification of appropriate policies. Once training is completed, the system is implemented by having the e-broker monitor those potential customers visiting the site, and adopting the policies selected by the users that are triggered by specific site visitor actions.

In person-to-person transactions, a broker negotiates a purchase or sale by acting as an intermediate agent between the two parties. The goal of the E-commerce broker presented here is to approximate the services that an actual broker or sales agent performs by detecting when a customer is losing interest, dynamically tailoring the delivery to accommodate the customer's needs, and facilitate the transaction.

The broker has two main tasks. First, it captures browser events that can lead to the determination of a likely customer departure. In order to do that, the broker intercepts the page delivered by the web server, inserts snippets of javascript code into it to be able to capture a set of predefined events, and redelivers the modified web page to the intended client. The second task of the browser is to analyze the browser requests. The broker intercepts the requests and usually forwards regular browser request to the server. However, in the presence of special "departure" requests generated automatically by the inserted code in the page, the broker retrieves the proper action for imminent departure from the policy database and sends the request to the server.

Data Collection

The intent of the proposed system is to apply web technology and consumer elicitation methods to better support an electronic commerce operation. Clickstream analysis is the activity by which a company or internet service provider can follow the trail of a web customer, to include pages visited, length of visit, etc. There are a number of companies providing clickstream analysis software, including ClickStream.com, netmetrics.com, and others that capture events at the client side. The system proposed here is envisaged as capable of doing this data collection dynamically, in response to specific web customer actions. In the proposed system, technology will be used to extract each consumer's past behavior in the form of actual purchases, and infer from this history a prediction of firm products that should be of interest. Classification and prediction of specific customer behaviors are modeled by (a) an adaptive decision theoretic model that can treat uncertainty either probabilistically or use other reasoning techniques based on rule-based systems; and (b) analysis of what each set of typical customers usually want (conjoint analysis).

The ideal pricing approach is to seek optimal pricing strategies, based on dynamic programming approaches (Leloup and Deveaux, 2001). Tellis and Zufryden (1995) presented this approach in the context we are discussing. The basic idea is to experimentally apply different price levels, gathering data on customer response. This approach involves an inherent tradeoff, in that the more experimental data gathered, the greater knowledge obtained, but the longer an optimal policy is delayed. Furthermore, identification of optimality in dynamic programming is well known to be intractable. For problems of even moderate size, dynamic programming may not be a viable approach in practice.

There have been attempts to apply machine learning to dynamic pricing on the internet (Kephart and Greenwald, 1999; Tesauro, 1999) sacrificing optimality for tractability. This approach is obviously limited in that it does not guarantee optimal decisions, but provides a way to gain improvements in results less limited by problem size.

Model

We present a formalization of the concept of the E-commerce broker. Variables defined are:

B_n is the observed behavior of customer n at the Internet web site. This is a random variable, measured and classified into categories such as 0 for non-buyer and 12 for buyer, or into m categories measuring visit periods at different web pages.

X_n is a binary variable. $X_n = 1$ if the customer buys, and $X_n = 0$ if the customer does not buy.

F_n is the forecast of probability $\Pr[X_n]$ given the observed behavior B_n. Based on how the customer behavior has been observed and partitioned, and what kind of data has been gathered, different forecasting methods could be employed. For example, one could use a Bayesian approach, or a rule based approach combining approximate reasoning procedures such as rough sets, fuzzy sets, or evidence theory. The result is an m by 2 matrix of probability forecasts where the columns are $X_n = 1$ and $X_n = 0$, and the rows are the customer categories.

D_n is the decision to offer incentives in the form of bonuses or additional services for the purpose of increasing the probability of purchase given the forecast F_n. Knowledge of the incremental impact of different online incentives on the probability of purchase is needed to determine the incremental reward matrix. This is what active analysis of non-invasive data needs to provide. The result is an m by K probability matrix where K is the number of different policies considered.

R_n is the amount of purchase by customer n, a random variable with distribution f(R), known after data gathering and analysis.

$C(d_n(f))$ is the cost of offering bonus d to customer n.

N is the number of customers at the site during the planning horizon. The objective is to maximize expected net return for the site.

Max Expected Return = $\displaystyle\sum_{n=1}^{N}\left[\left(\int \Pr(X=1 \perp f_n, d_n(f))R_n f_n(r)dR\right) - C(d_n(f))\right]$

Subject to any limitations on resources needed to offer the bonuses.

Tellis and Zufryden (1995) presented a similar model in terms of dealing with the question of which brands to discount, by how much, and when.

We propose a framework intended to rank customers on their estimated return given past customer behavior and historical data. This framework includes identification of site management policies to maximize expected return.

Customer behavior is a random variable affecting the outcomes of the customer's visit, which is also a random variable. Let $\Pr[X_n = 1] = p_n$. We can assume that p_n is constant at least over a short period (which assumes customer static behavior). Customers can be classified based on their likelihood of making a purchase in any visit. Let m be the number of states in the classification. Then m depends on the observation capabilities of the system.

Observations of customer actions and historical data can be used to dynamically forecast customer state X_n. This forecast can be deterministic, specifying one state for the customer in each time period, or probabilistic, giving the probability of customer association with each state. Forecast outcomes will be used to make decisions in terms of offering a discount or changing content. The decision is made based on policies, which are the rules mapping customer state to a site reaction.

The decision of the system will affect customer state and the probability of purchase. Final estimation of purchase probability and the costs of the decision will determine return R_n, which is a random variable with pdf $f(.)$.

The broker must be able to monitor all communication between the web server and the client browser. It could be located in the client, in the server, or between the two. We have chosen to implement the broker as a software appliance that executes on a device located between the server and the client, as shown in Figure 2. Our choice for implementing the broker in this manner is based on the desire to be independent of the server technology. For example, in our implementation, the choice of the server software, the application server technology, and even the database server has no affect on the broker, which increases the overall transparency.

Conclusions and Future Research

E-commerce involves the opportunity to better deliver material over electronic markets to customers. Part of the process of this development is generation of new and better techniques for dealing with business problems. Negotiating in a retail environment (haggling) is something that in the past has been missing from E-commerce. That may have been a good thing. However, the potential to detect loss of customer interest in purchasing provides additional opportunity for E-commerce sites to increase their sales.

This paper presents a decision support concept of how E-commerce retailers can gather individual customer preference information, and monitor customer behavior and intelligently select responses aimed at optimizing profitability. A framework was developed formalizing the type of data required to forecast customer behavior for each of the available set of policy actions.

Preference information can be captured through initial and follow-up surveys. Capture of customer web behavior is through on-line monitoring of events at the client web pages visited by the client. Mouse coordinates can be monitored to keep track of the activities that in the past have led to departure of the web customer without purchasing. Thus there are components of data collection through clickstream analysis, conjoint analysis (in the sense of forecasting web customer final action given specific web actions) to identify the best policy (in the sense of optimizing profit), and expert systems to implement policies identified in the conjoint analysis phase. The effectiveness of policies developed on the basis of this first form of data collection will be compared with a more complete and invasive elicitation of what web customers in given states are likely to do is proposed through a form of conjoint analysis. Input from conventional survey-based conjoint preference models will be compared with preference models inferred from clickstream generated input.

The continuation of this project will implement a proof-of-concept E-commerce broker with an emphasis on mechanism over policy. Data can be acquired actively or through non-invasive methods. The research will examine the tradeoff between level of detail gathered and system response time

References

Berthon, P., Pitt, L.F., and Watson, R.T., "The World Wide Web as an Advertising Medium," Journal of Advertising Research (36:1), 1996, p-p. 43-54.

Clickstream Audience Measurement Solutions. http://www.clickstream.com/

Cooley, R., Mobasher, B., and Srivastava, J., "Data Preparation for Mining World Wide Web Browsing Patterns. Journal of Knowledge and Information Systems (1:1), 1999.

DeSarbo, W.S., Jedidi, K. and Sinha, I., "Customer Value Analysis in a Heterogeneous Market," Strategic Management Journal (22), 2001, 845-857.

Dreze, X. and Zufryden, F., "Testing Web Site Design and Promotional Content," Journal of Advertising Research (37:2), 1997, 77-91.

Dreze, X. and Zufryden, F., "Is Internet Advertising Ready for Prime Time?" Journal of Advertising Research (38:3), 1998, 7-18.

Eighmey, J., "Profiling User Responses to Commercial Web Sites," Journal of Advertising Research (37:2), 1997, 59-66.

Fraenkel, L., Bodardus, S., and Wittink, D.R., "Understanding Patient Preferences for the Treatment of Lupus Nephritis with Adaptive Conjoint Analysis," Medical Care (39:11), 2001, 1203-1216.

Green, P.E., Krieger, A.M. and Wind, Y., "Thirty Years of Conjoint Analysis: Reflections and Prospects," Interfaces (31:3), May-June 2001, S56-S73.

Green, P.E. and Srinivasan, V., "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice," Journal of Marketing (54), 1990, 3-17.

Gustafsson, A., Herrmann, A., and Huber, F., Conjoint Measurement: Methods and Applications, 2nd ed., New York: Springer, 2001.

Kangas, S. and Chiu, A. Learning from Event Logging, www.netconversions.com, 2001.

Kephart, J. and Greenwald, A.,"Shopbot and Pricebots," Proceedings of the International Joint Conference on Artificial Intelligence, 1999.

Kumar, K. and van Hillgersberg, "ERP Experiences and Evolution," Communications of the ACM (43:4) April 2000, 23-26.

Leloup, B. and Deveaux, L., "Dynamic Pricing on the Internet: Theory and Simulations," Electronic Commerce Research (1:3), 2001, 265-276.

Mobasher, B., Cooley, R., and Srivastava, J. "Automatic Personalization Based on Web Usage Mining," Communications of the ACM (43:8), August 2000, 142-151.

Montgomery, A.L. "Applying Quantitative Marketing Techniques to the Internet," Interfaces (31:2), March-April 2001, 90-108.

Spiliopoulou, M., Pohle, C., and Faulstich, L.C., "Improving the Effectiveness of a Web Site with Web Usage Mining," In KDD'99 Workshop on Web Usage Analysis and User Profiling WEBKDD'99, B. Masand and M. Spiliopoulou, eds., San Diego, CA: ACM, 1999.

Tellis, G.J. and Zufryden, F.S., "Tackling the Retailer Decision Maze: Which Brands to Discount, How Much, When and Why?" Marketing Science (14:3, part 1), 1995, 271-299.

Tesauro, G., "Pricing in Agent Economies Using Neural Networks and Multi-Agent Q-Learning," Proceedings of Workshop Learning About, Form and With other Agents, 1999.

Van Everdingen, Y., van Hillegersberg, J. & Waarts, E. "ERP Adoption by European Midsize Companies," Communications of the ACM (43:4), April 2000, 27-31.

Wittink, D.R. and Cattin, P., "Commercial Use of Conjoint Analysis: An Update," Journal of Marketing (53), 1998, 91-96.