MARKETING SCIENCE INSTITUTE

MSI Reports

2005

WORKING
PAPER
SERIES

ISSUE ONE

NO. 05-001

MARKETING SCIENCE INSTITUTE

MSI Reports

Executive Director
Leigh M. McAlister

Research Director
Ross Rizley

Editorial Director
Susan Keane

Publication Design
Laughlin/Winkler, Inc.

2005

WORKING
PAPER
SERIES

ISSUE ONE

NO. 05-001

The Marketing Science Institute supports academic research for the development—and practical translation—of leading-edge marketing knowledge on issues of importance to business performance. Topics are identified by the Board of Trustees, which represents MSI member corporations and the academic community. MSI supports academic studies on these issues and disseminates findings through conferences and workshops, as well as through its publications series.

Marketing Science Institute
1000 Massachusetts Avenue
Cambridge, MA
02138-5396

Phone: 617.491.2060
Fax: 617.491.2065
www.msi.org

MSI Reports (ISSN 1545-5041) is published quarterly by the Marketing Science Institute.

Working Paper Series
The articles that appear in *MSI Reports* have not undergone a formal academic review. They are released as part of the MSI Working Paper Series, and are distributed for the benefit of MSI corporate and academic members and the general public.

Subscriptions
Annual subscriptions to *MSI Reports* can be placed online at www.msi.org. Questions regarding subscriptions may be directed to pubs@msi.org.

Single reports
Articles in *MSI Reports* are available in downloadable (PDF) format at www.msi.org.

Past reports
MSI working papers published before 2003 are available as individual hard-copy reports; many are also available in downloadable (PDF) format. To order, go to www.msi.org.

Corporate members
MSI member company personnel receive all MSI reports (PDF and print versions) free of charge.

Academic members
Academics may qualify for free access to PDF (downloadable) versions of MSI reports and for special rates on other MSI print publications. For more information and to apply, go to "Qualify for academic membership" on www.msi.org.

Classroom use
Upon written request, MSI working papers may be copied for one-time classroom use free of charge. Please contact MSI to obtain permission.

Search for publications
See the searchable publications database at www.msi.org.

Submissions
MSI will consider a paper for inclusion in *MSI Reports*, even if the research was not originally supported by MSI, if the paper deals with a priority subject, represents a significant advance over existing literature, and has not been widely disseminated elsewhere. Only submissions from faculty members or doctoral students working with faculty advisors will be considered. "MSI Working Paper Guidelines" and "MSI 2004-2006 Research Priorities" are available in the Research section of www.msi.org.

Publication announcements
To sign up to receive MSI's electronic newsletter, go to www.msi.org.

Change of address
Send old and new address to pubs@msi.org.

Offering Online Recommendations to Impatient First-time Customers with Conjoint-based Segmentation Trees

Arnaud De Bruyn, John C. Liechty, Eelko K.R.E. Huizingh, and Gary L. Lilien

Can websites offer relevant recommendations to their online visitors quickly and without extensive consumer inputs? In this study, the authors leverage conjoint analysis techniques to design a recommender system that uncovers customers' preferences based on demographics and product usage.

Arnaud De Bruyn is Assistant Professor of Marketing, ESSEC Business School, Cergy-Pontoise, France.
John C. Liechty is Assistant Professor of Marketing and Statistics, The Pennsylvania State University.
Eelko K.R.E. Huizingh is Associate Professor of Marketing and Information, University of Groningen, The Netherlands.
Gary L. Lilien is Distinguished Research Professor of Management Science, The Pennsylvania State University.

Report Summary

Online consumers are known to be impatient. They want to make their purchases in as little time as possible. In order to maximize the revenue generated by these customers, merchant websites must be able to offer relevant recommendations quickly and with minimal customer input. Such recommendations are particularly hard to make to first-time visitors since no information exists on their preferences and purchase patterns. In this report, De Bruyn, Liechty, Huizingh, and Lilien propose a framework to help companies develop recommendation systems that can provide quality advice to their impatient and first-time online visitors.

Typically, firms have relied on conjoint analysis to assign customers to product preference groups for the purpose of marketing. De Bruyn, Liechty, Huizingh, and Lilien demonstrate how this technique can be leveraged to design a recommender system that uncovers customers' preferences based on simple demographics and product usage questions rather than the extensive inputs such an analysis usually requires.

To begin, they perform a conjoint analysis on a representative sample of individuals to elicit and measure their preferences. In addition to rating a set of products, they are also asked to answer demographic and product usage questions. The authors compare three approaches for linking individuals' characteristics to their preferences in order to identify the most informative demographic and product usage questions.

They found that the stepwise componential regression method allowed them to develop a sequence of questions that predicted customer preferences more efficiently and accurately than the two other approaches (cluster-classification and Bayesian treed regression).

For managers, developing such a recommendation system would allow merchant websites to satisfy both their own need to direct customer purchases and their customers' desire to move quickly through the interaction. ■

Introduction

"Today's online consumers are more impatient than ever before" —*Michael Silverstein, BCG's Consumer Practice group*

"Rookies are the most impatient visitors of all" —*Real Business*

Online consumers are impatient (De Angelis 2001; Enos 2001), and it is crucial for e-commerce websites to offer relevant recommendations to their online visitors quickly and with minimal customer burden. This is especially important with new visitors who are often unwilling to engage in time-consuming preference elicitation procedure. But most existing systems that offer recommendations to online visitors are poorly suited to make recommendations to new visitors for whom no prior information is available and who may lack product-category expertise.

Collaborative filtering recommender systems (Resnick and Varian 1997; Schafer, Konstan, and Riedl 2001) would appear to provide a solution. They are "agents that use behavioral or preference information to filter alternatives and make suggestions to a user" (Ansari, Essegaier, and Kohli 2000). But such systems require prior information about visitors, such as product ratings or implicit preferences inferred from browsing behaviors or purchase history, and are therefore impractical for first-time visitors.

Consumer-decision support systems (CDSS) potentially offer alternative solutions. A CDSS is "a system that connects a company to its existing or potential customers, providing support for some part of the customer decision-making process" (O'Keefe and McEachern 1998). Early developers assumed that by facilitating the exploitation of information and expanding processing capabilities, users of CDSS's were likely to compare more alternatives, evaluate them more completely, and thus make better decisions (Hoch and Schkade 1996). Most CDSS's assume that customers are willing and capable of comparing alternatives on those performance dimensions that are relevant and important to them. This assumption is questionable with complex, intangible, or highly customizable products as well as when customer expertise is low (Grenci and Todd 2002; Huffman and Kahn 1998) or when potential customers are not highly involved in the decision, impatient, and unwilling to go through a time-consuming evaluation procedure. Under such circumstances, it can be risky for a commercial website to count on their visitors' motivation–and patience–to use a CDSS to find the right product for them.

The objective in this paper is to develop and assess the performance of a method to offer personalized recommendations to online visitors under the following constraints:

- The method should not require any prior knowledge about the consumer
- Consumer input should be minimal
- Product-category expertise should not be needed to use the recommender system

We first present three competing methods that use the results of a conjoint study conducted ex-ante to develop an optimal sequence of questions and elicit customers' preferences. We then report the results of an empirical study in which each method was tested. We conclude by a discussion of the results and their managerial implications.

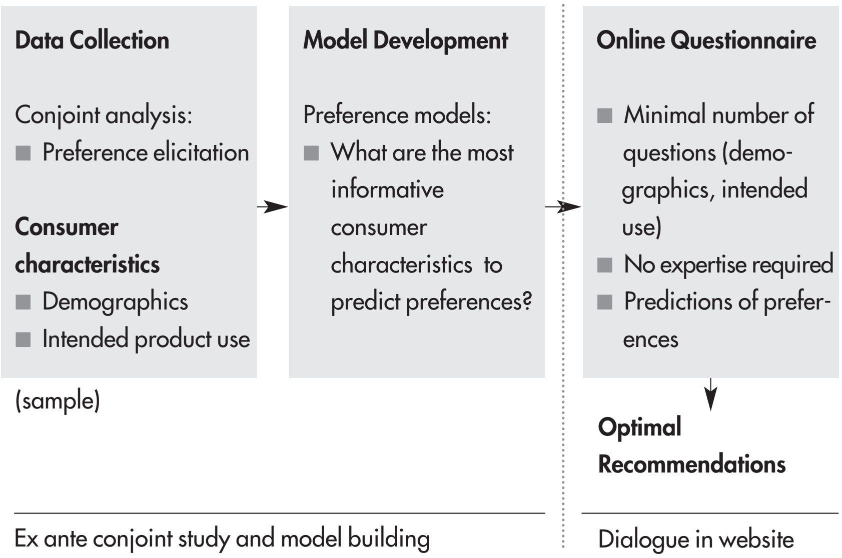
Methodology

Ansari, Essegaier, and Kohli (2000) suggest that preference models used in marketing, such as conjoint analysis, offer good alternatives to collaborative and information filtering recommender systems when prior behavioral data about an individual is sparse. Despite significant efforts to make preference elicitation procedures quicker and more efficient (Sawtooth 1991; Toubia, Simester, Hauser, and Dahan 2003), conjoint analysis still requires considerable consumer input and is impractical for use as a

Figure 1

Conjoint-based Recommender System

We propose a three-step approach to develop an optimal sequence of questions to elicit online visitors' preferences and make optimal recommendations based on a conjoint study conducted beforehand.



recommender system. Nevertheless, conjoint analysis offers a useful way to model consumers' preferences. Hence, we explore the possibility that a conjoint study conducted beforehand on a sample of customers can be leveraged to design a recommender system capable of offering personal recommendations to online visitors with minimal inputs. We employ the following approach (See Figure 1):

- We perform a conjoint analysis on a representative sample of individuals who, in addition to rating a set of products, are also asked to answer demographic and product usage questions
- We link respondents' characteristics to their preferences, and we identify the most informative demographic and product usage questions
- We use the results of the previous analysis to develop an optimal sequence of questions to elicit consumers' preferences

We now discuss three competing methods to link individuals' responses to their preferences in a way that can be operationalized in an online questionnaire.

Cluster-classification

A natural approach to this problem is to follow a three-step, estimation-clustering-classification approach, similar to those commonly implemented to segment and target customers in direct marketing. After conducting the conjoint study and collecting individual-level responses (e.g., ratings, preferred choices, or pairwise comparisons of conjoint profiles), (1) individuals' preference partworths are estimated with standard estimation procedures; (2) respondents are then clustered into segments of similar needs and preferences using either hierarchical or nonhierarchical methods (Green and Krieger 1991); and finally (3) descriptor variables, or segmentation bases (e.g., demographics characteristics, intended product usage) are used to predict segment membership of each individual. We will refer to these three steps as the *estimation*, *clustering*, and *classification* stages respectively.

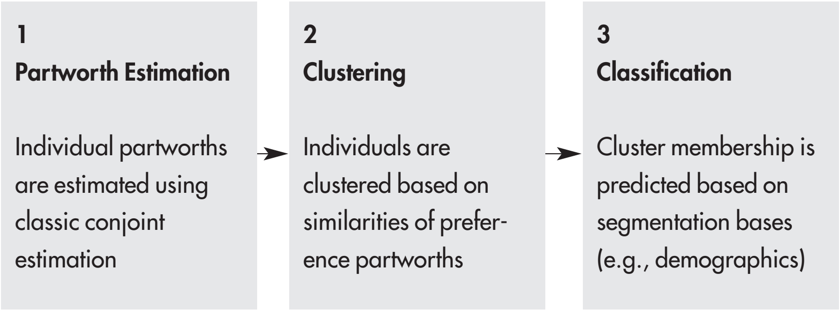
Translated into our context of recommender system, the classification procedure would point out the best questions to ask in order to determine to which segment of preferences an online visitor is most likely to belong. Predicted segment membership could then be exploited to make recommendations best suited for a typical member of the identified segment.

The estimation, clustering, and classification procedures are usually envisaged separately in academic (Green and Krieger 1991) and commercial applications (Wittink, Vriens, and Burhenne 1994), but this approach has the major inconvenience of grouping respondents based on possibly unreliable individual-level estimates (the degrees of freedom at the individual level being usually rather small.) In addition, conjoint models that are overparameterized at the individual level cannot be accommodated: respondents have to rate at least as many profiles as there are attribute levels to be estimated. Finally, since segments are formed regardless of how well they can be separated for

Figure 2

Traditional Conjoint-based Segmentation

Traditional conjoint-based segmentation usually follows a three-stage procedure: (1) individual preference partworths are estimated for each respondent, (2) individuals are grouped into homogeneous segments based on preference partworth similarities (e.g., K-Means), and (3) segment membership is predicted based on available descriptors (e.g., discriminant analysis). Researchers have argued that grouping two or more steps together could increase overall performance.



1
Partworth Estimation

Individual partworths are estimated using classic conjoint estimation

2
Clustering

Individuals are clustered based on similarities of preference partworths

3
Classification

Cluster membership is predicted based on segmentation bases (e.g., demographics)

For methods grouping steps 1 and 2, see Hagerty (1985), Kamakura (1988), DeSarbo, Oliver, and Rangaswamy (1989), Wedel and Kistmaker (1989), Wedel and Steenkamp (1989), DeSarbo, Wedel, and Vriens (1992)

For methods grouping steps 1, 2, and 3 all together, see Gupta and Chintagunta (1994), Kamakura, Wedel, and Agrawal (1994), Wedel and Steenkamp (1991)

targeting purposes, clustering and classification stages might achieve suboptimal solutions when considered separately since they try to maximize two separate objective functions independently.

It has been suggested that traditional conjoint-based segmentation could be improved by grouping two or more stages together. Figure 2 summarizes these developments.

Researchers have developed various methods to group the estimation and clustering stages into a one-step procedure that optimizes a single objective function. Such methods include Q-Factor analysis (Hagerty 1985), hierarchical clustering (Kamakura 1988), clusterwise regression (DeSarbo, Oliver, and Rangaswamy 1989; Wedel and Kistmaker 1989; Wedel and Steenkamp 1989), and mixture regression methods (DeSarbo, Wedel, and Vriends 1992). Although the latter approach seems to offer the most promising results, a Monté Carlo study (Vriens, Wedel, and Wilms 1996) showed that, in term of predictive accuracy, no method outperforms the traditional two-step approach. This finding can be explained by the within-segment heterogeneity that affects all methods (Wedel and Kamakura 2000) and flattens out their performances at very similar levels. As a result, although the above methods may greatly differ on other performance criteria (Vriens, Wedel, and Wilms. 1996), they do not significantly affect the metrics we are interested in, namely the predictive accuracy of preference partworths based on segment membership.

Other academics have proposed methodologies to group the estimation, clustering, and classification stages together in an integrated framework (Gupta and Chintahunta 1994; Kamakura, Wedel, and Agrawal 1994; Wedel and Steenkamp 1991). These algorithms have great merits, but are not suited for the application at hand primarily because they draw *simultaneously* (as opposed to s*equentially)* on all available descriptors to assign segment membership. As a result, these methods cannot point out the most informative questions to ask and do not indicate in which optimal order bits of information should be gathered to predict segment membership.

The same problem arises with traditional classification methods such as discriminant analysis, artificial neural network, or multiple regressions on indicator matrix. These methods use all available information at once and are, therefore, not suited (or at least not specifically designed) to identify a sequence of questions to predict to which cluster of preferences an individual is most likely to belong.

A variant of discriminant analysis, namely stepwise discriminant analysis, might offer the promise of a solution. But this method has been severely criticized in the literature, and although it might serve as a useful exploratory tool, it has little value for a holdout sample predictive application (Huberty 1994; Huberty 1989).

The CART (Classification and Regression Trees) algorithm offers yet another interesting alternative (Breiman et al. 1984). This method sequentially splits a population (the parent node) into child nodes, such that each child node is populated with individuals as pure as possible in terms of class membership. Then, each child node becomes a parent node itself, and is subsequently split, and so on, until a stopping rule is reached. The tree is eventually pruned back, based on a cost-complexity criterion, to reduce overfitting and enhance holdout sample predictive accuracy. Ideally, each end node of the tree becomes perfectly pure: it only contains individuals from one and only one segment and thus achieves a perfect classification.

The CART algorithm has several interesting properties for the application at hand. First, the tree structure conveyed in the solution can readily be translated into an optimal sequence of questions, each split pointing out the best additional bit of information to refine prediction of segment membership. Second, two child nodes could be expanded using two different splitting rules (i.e., subsequent best splits might differ in the left and right child nodes), hence, splits are locally optimal. Third, it is straightforward enough to make recommendations based on the tree's structure: preferences of the segments populating each node can be translated into optimal recommendations for a holdout sample population.

For this research, we integrate the above considerations and embrace the traditional, three-step approach as follows:

- Estimation: Individual-level preference partworths are first estimated using classical conjoint equations
- Clustering: Individual partworth estimates are then clustered into preference segments (We favor a nonhierarchical clustering methodology, i.e., K-Means, because centroids have an immediate interpretation, namely they represent the average preference partworths of the segment's population and can be readily translated into optimal recommendations)
- Classification: Cluster membership is predicted based on descriptor variables using CART. The tree structure conveys both the sequence of questions to ask (i.e., the sequence of descriptor variables employed by the CART algorithm to split the population) and optimal recommendations to make (i.e., average preference partworths of each node's population)

This methodological choice has the advantage of being easily replicable with standard statistical packages, which makes it more likely to be implemented by practitioners.

A note on integrated approaches

Several researchers have claimed that segmenting and pooling "similar" individuals could improve prediction for each individual in general (e.g., Bucklin and Gupta 1992) and in conjoint analysis in particular (DeSarbo et al. 2002; DeSarbo, Oliver, and Rangaswamy 1989; DeSarbo, Wedel, and Vriens 1992; Green, Kriefer, and Schaffer 1993; Kamakura 1988; Ogawa 1987). To pool individuals increases the degrees of freedom of the model, leads to more stable and accurate partworth estimates, and prevents the model from overfitting individual-level data.

The two solutions we will introduce, Bayesian treed regression and stepwise componential segmentation, follow a similar line of thinking. They both integrate estimation, clustering, and classification into a unique stage and pool data obtained from "similar" individuals to generate more reliable preference partworth estimates. Unlike the one-stage methods cited earlier, classification is achieved sequentially and hence is suited for the identification of an optimal sequence of questions.

Bayesian treed regression

The idea behind *Bayesian treed regression* is to partition a dataset using a tree structure, but instead of computing a simple mean or proportion, to fit a different regression model at each end node (Chipman, George, and McCulloch 2002). It is in a sense a combination of CART and clusterwise regression, implemented within a Hierarchical Bayes regression framework.

Although Chipman, George, and McCulloch (2002) did not develop the algorithm with conjoint analysis in mind, its application to this domain is straightforward. The treed regression algorithm simultaneously (1) clusters individuals into nonoverlapping segments (i.e., nodes) through a tree structure, (2) pools profile ratings made by individuals in the same node into a unique regression model, and (3) estimates preference partworths using Hierarchical Bayes regression. Since parameters of the conjoint model are computed at the node-level, treed regression eludes the overfitting issue of individual-level models (unless, of course, a single individual happens to populate a node). Furthermore, segments are formed on the basis of a sequence of binary splits performed on descriptor variables. The segments, therefore, are perfectly identifiable, and, similarly to the CART algorithm, the optimal sequence of questions is naturally embedded in the solution.

Note that, in contrast to traditional tree methods that apply locally optimal, greedy splitting rules, Bayesian treed regression tries to achieve global optimality by searching the space of possible models using MCMC (Markov Chain Monte Carlo) exploration.

Stepwise componential segmentation

While both CART and Bayesian treed regression are natural candidates for the design of an optimal sequence of questions (the successive splits signal what questions to ask and in what order, while the tree structure permits the system to choose the next best question based on the customer's previous answers), they may be subject to problems of overfitting. Because the population is successively divided, each decision to further split the population involves a smaller portion of the dataset, possibly leading to overfitting. In addition, data requirements grow exponentially with tree size.

In the past, several hybrid methods have been proposed to overcome the heavy data requirements of classic conjoint analysis and to reduce data collection effort and time (Green 1984). One of these approaches, *componential segmentation* (Green and DeSarbo 1979), explicitly incorporates respondent descriptor variables in the utility function by re-expressing individuals' preference partworths as linear combinations of descriptor variables.

We use the following notations:

$1..i..I$	refers to respondents
$1..j..J$	refers to profiles rated by each respondent
$1..k..K$	refers to preference partworths to be estimated by the model, i.e., one per attribute level (excluding all dummy levels set to 0 for identification purpose), including an intercept
$1..q..Q$	refers to respondents' descriptor variables, such as demographics, intended product usage, etc.
y_{ij}	is the preference score given by the i^{th} individual to the j^{th} profile
β_i	is the vector of preference partworth of the i^{th} individual (K elements)
P_{ij}	is the vector of attribute levels of the j^{th} profile rated by the i^{th} individual (K elements)
D_i	is the vector of descriptor variables pertaining to the i^{th} individual (Q elements)
Ψ	is a matrix of parameters to be estimated (K rows and Q columns), and is not specific to a particular individual

We build a traditional conjoint model where *predicted* preference scores $\tilde{y}$ are linear combinations of preference partworths and attribute levels, $\tilde{y}_{ij} = (\beta_i \cdot P_{ij})$, such that they minimize $SSE = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - \tilde{y}_{ij})^2$. While in a classic conjoint model, all vectors β_i are individually computed, we re-express $\beta_i = (\psi \cdot D_i)$: $\forall i$ and optimize Ψ at the population level. Parameters to be estimated go from $I.K$ in traditional conjoint to $Q.K$ in componential segmentation, with Q commonly much smaller than I, hence increasing the degrees of freedom of the model.

Although this approach is very natural (e.g., it seems reasonable to expect price sensitivity to be a function of income, or preferences for specific benefits to be linked to demographics, lifestyle characteristics, or intended product usage), componential segmentation has had a limited impact in practice. The reasons are twofold. First, componential segmentation "leads only to subgroup utility functions because all respondents with a similar background profile are assumed to have the same utility function" (Green 1984, p. 156). In other words, if $D_i = D_j$, then $\beta_i = \beta_j$; if two respondents have identical descriptors, these individuals are assumed to have identical preferences, too.

Second, the success of this method depends on the existing correlations between consumers' characteristics and individual preferences (Wedel and Kamakura 2000). If the latter are loosely related to observable respondents' characteristics, i.e., if descriptors are not good predictors of the 'true' preference partworths, results will be disappointing. Given the nature and limited amount of information usually available in segmentation applications, this can be a serious impediment.

This limitation, however, is much less critical in a recommender system context, since very specific questions can be asked to online visitors, including questions pertaining to usually unobservable consumer characteristics (i.e., needs, likes and dislikes, experience), broadening the range of possible and relevant questions.

In its original format, all information is included in the componential segmentation model, and preference partworths' estimation draws on all Q individuals' descriptors. In order to determine an optimal sequence of questions, we need to adapt the original algorithm into a stepwise procedure.

We proceed as follows. First, we begin by assuming that $Q = 1$ (vector of descriptors is of size 1), and set $D_i = \{1\}$: $\forall i$ (equivalent to an intercept in linear regression). Hence, Ψ is a vector of size K, and $\beta_i = \Psi$: $\forall i$. Ψ is equivalent to the average preference partworths of the population as a whole, which minimizes SSE.

The vectors of descriptor variables are then augmented by one element ($Q' \leftarrow Q + 1$) at each step of the stepwise procedure. The next descriptor included in the model is the one that minimizes SSE, conditional on the optimization of the new matrix Ψ. As with the CART algorithm, the selection of the next most informative descriptor is achieved by testing all possible descriptors one by one[1] as potential candidates to fill the last element of D, and by eventually adding in the descriptor matrix the one that leads to the highest incremental improvement. At each step, matrix Ψ is augmented by one column, as is the number of parameters to be estimated by K elements.

A "statistically optimal" stopping rule would be to test the hypothesis that the last K parameters added to the model are zero. If this is true, the new model is no better than the previous one with K fewer parameters, the last descriptor is removed, and the stepwise procedure stops. This hypothesis can be tested with an F-test, distributed as $F_{K, IJ-K(Q-1)}$ (Rencher 1995, p. 359).

In practice, however, the large number of degrees of freedom will make the rejection of the null hypothesis very unlikely (IJ represents the total number of profiles rated by all respondents—several thousands in most conjoint studies). Translated into our context of recommender systems, if this stopping rule were to be applied, the questionnaire suggested by this method would be too lengthy. The recommender system would keep asking for additional information as long as it made "statistical" sense, even if improvements in actual recommendations were marginal.

Table 1
Attributes and Attribute Levels of the Research Design

Most respondents prefer a five-day weather report, general university news, a $4 online coupon, and a mix of both general and business news. The most important attribute in explaining preferences for particular pages is the general news attribute.

Attribute	Levels	Preferred level	Variance explained
Weather report	5-day forecast	60%	15%
	1-day extended report	40%	
University news	General news	51%	17%
	Sports news	49%	
Online coupon	$2	26%	12%
	$4	74%	
General news	U.S. news only	11%	29%
	Mix U.S./world	76%	
	World news only	13%	
Business news	Stocks news only	8%	27%
	Mix stocks/general	58%	
	General news only	34%	

We choose a more practical rule and stop the development of the questionnaire when the inclusion of an additional descriptor does not improve the adjusted R^2 (between y and $\hat{y}$) by at least .005.

In contrast to the optimal sequence of questions suggested by the two previous tree-based methods, stepwise componential regression generates static questionnaires: all online visitors get the same questions in the same order, independently of their answers to previous questions. Each question selection, however, is optimized on the entire dataset, which is likely to enhance the robustness of the method and reduce the risk of overfitting.

Empirical Study

Research design

We asked 616 graduate and undergraduate students at a large, northeastern U.S. university to rate customized web pages from a hypothetical university news portal. The pages were described by five attributes: weather report, university-related news, general news, business news, and value of an online coupon. Attributes had either two or three levels. Table 1 reports the five attributes and their different levels, as well as estimated average preferences in terms of preferred levels and variance explained, estimated using individual-level conjoint models.

We conducted the study electronically in a controlled lab setting, and the pages to be rated were displayed on participants' computer screen. The study comprised four sections: (1) a first task, designed to familiarize respondents with the attributes and the software system used to collect the data; (2) a conjoint task that asked respondents to rate 21 web pages, displayed one at a time on the screen, on a 100-point preference scale; (3) a self-administered questionnaire, with 99 questions concerning socio-demographics, consumption habits, likes and dislikes; and (4) a holdout task that asked respondents to distribute 100 points among 4 different news pages. All participants repeated the holdout task five times with different sets of pages in each replication.

In terms of the selection of the 21 profiles showed to participants during the main conjoint task (2), we built four partially balanced blocks using orthogonal fractional factorial design, and randomly assigned each participant to one of these blocks. The order of the profiles to be rated was randomly rotated within each block and across respondents to avoid presentation effect.

The questions in section 3 were selected as the result of a pilot study in which 43 students, with backgrounds similar to those of the core sample, were asked to explain the rationale behind their preferences for certain attributes. Experts suggested additional questions. For instance, one of the questions retained was whether or not participants owned stocks, a likely influence on their preferences for business news. Of the 99 questions, 44 were dichotomous, 7 were multiple-choice questions, and 48 were Likert items.

Table 2

Average Preference Partworths of the Five Identified Clusters of Respondents

A typical member of the third cluster highly values sports news and U.S. news, and does not care about the face value of the coupon or the type of business news displayed. All types of general news are okay as long as they contain U.S. news. Note that a mix of U.S. and general world news seem to dominate systematically the other news options within each cluster.

Attribute	**Levels**	**C1** *N = 174*	**C2** *N = 92*	**C3** *N = 61*	**C4** *N = 86*	**C5** *N = 103*
Intercept		11	53	39	40	32
Weather report	5-day forecast (*)					
	1-day extended report	2	–7	12	0	–21
University news	General news (*)					
	Sports news	3	–17	18	–25	15
Online coupon	$2 (*)					
	$4	12	1	2	9	10
General news	U.S. news only (*)					
	Mix U.S./world	30	5	4	26	17
	World news only	12	–17	–20	14	3
Business news	Stocks news only (*)					
	Mix stocks/general	28	19	4	6	17
	General news only	17	20	0	9	6

(*) Dummy levels set to zero.

In the last task (4), all respondents distributed 100 points among the five same sets of four pages (to increase comparability), but the order of presentation was randomly rotated.

Analysis

We randomly split the data into a calibration set ($N = 516$) and a testing set ($N = 100$). For the cluster-classification algorithm, after estimating and scaling the individuals' partworths, we grouped respondents into clusters of preferences using the K-Means algorithm and found that five groups worked best. Increasing the number of clusters beyond five did not enhance the performance of the cluster-classification method in terms of holdout sample fit. Table 2 reports the size and average preference partworths of the five identified clusters (numbered C1 to C5).

We grew the classification tree using the CART algorithm, stopped the splitting process when we reached a minimum node size, and then pruned back the tree using the cost-complexity criterion (Breiman et al. 1984). The final tree contained 20 end nodes with a maximum depth of 9 splits.

For the Bayesian treed algorithm, given our objective to require minimal consumer input, we set the four parameters that govern the splitting decisions to values that favor small trees ($\alpha = .5$, $\beta = 2$, $c = 1$ and $\lambda = .404$; see Chipman, George, and McCulloch [2002] for discussion). We ran a large number of iterations to assure the stability of the solution, and we dedicated one-third of the calibration set to internal overfitting diagnostic. The final tree contained 23 end nodes with a maximum depth of six splits.

The stepwise componential regression did not require any parameterization, and tests of the modifications in the adjusted R^2 led us to stop the development of the model after the second question.

Table 3
Example of Questionnaire

Example of questionnaire suggested by the cluster-classification method, and associated responses from the 53rd respondent (randomly selected). Prior to the first question, the probability of an individual to belong to the third cluster is equal to the proportion of this cluster in the sample, i.e., 11.8%. After five questions, the system estimates that this particular individual has 85.7% chance to belong to the third cluster and stops asking additional questions. Recommendations are then optimized based on the prediction of cluster membership. Each question corresponds to a node in the tree developed by the CART algorithm.

Question	**Answer**	**Probability to belong to cluster 3**
Initial proportion (size of the cluster in parent node)		11.8%
Q1. General [university] news is typically more important to me than [university] sport news.	"No"	21.7%
Q2. With regards to local weather reports, detailed summaries of today's weather are typically more important to me than a less detailed five-day forecast.	"Agree"	39.3%
Q3. In the last month how many times have you eaten at a restaurant? (do not include on campus restaurants)	"3 to 6 times"	50.0%
Q4. Have you ever taken a seminar or class about the Web or Internet?	"No"	63.3%
Q5. How many men's home basketball games have you attended so far this season?	"3"	85.7%

Table 4
Stepwise Componential Segmentation's Preference Estimates after *N* Questions

Estimated elements of matrix Ψ. Elements in grey are not significant at $p < .05$.

Step	**Descriptors**	**Vector of preference partworths (*)**							
		Intercept	1-day extended report	Sports news	$4 coupon	Mix U.S./world	World news only	Mixed stocks/ general	General news only
	Base	45.4	–7.2	–2.9	3.6	11.8	–8.1	13.5	8.2
Q1	Base	40.4	–8.1	13.4	4.0	10.6	–9.1	10.9	5.8
	Descriptor 1 (a)	+7.5	+1.3	–24.7	–0.6	+1.9	+1.6	+4.0	+3.6
Q2	Base	33.5	7.5	14.0	4.3	10.0	–9.8	9.3	5.1
	Descriptor 1	+7.7	+0.9	–24.7	+0.6	+1.9	+1.5	+4.0	+3.6
	Descriptor 2 (b)	+10.7	–24.3	–0.9	–0.6	+1.0	+1.3	+2.5	+1.0
Q3	Base	38.3	4.8	15.4	3.8	10.8	–9.4	8.1	–3.7
	Descriptor 1	+8.9	+0.1	–24.2	–0.7	+2.6	+2.0	+3.6	+0.8
	Descriptor 2	+11.0	–24.5	–0.8	–0.5	+0.9	+1.5	+2.4	+0.5
	Descriptor 3 (c)	–7.5	+4.4	–2.3	+0.9	–1.6	–1.1	+2.1	+14.6

(*) The following attribute levels are set to 0 for identification purpose: five-day forecast (weather report attribute), general news (university news), $2 coupon (online coupon), U.S. news only (general news), stocks news only (business news).

(a) Descriptor 1: Update as indicated if respondent answers "yes" to the question "General [university] news is typically more important to me than [university] sport news."

(b) Descriptor 2: Update as indicated if respondent answers between 4 ("Disagree") and 6 ("Strongly disagree') to the question "With regards to local weather reports, detailed summaries of today's weather are typically more important to me than a less detailed five-day forecast."

(c) Descriptor 3: Update as indicated if respondent answers "yes" to the question "General business news is typically more important to me than stock market news."

Table 5

Comparison of the Three Methods to Full-Profile Conjoint Analysis

The stepwise componential regression method dominates all the other methods, including the classic, full-profile conjoint analysis, both in terms of efficiency (number of questions) and holdout sample predictive accuracy.

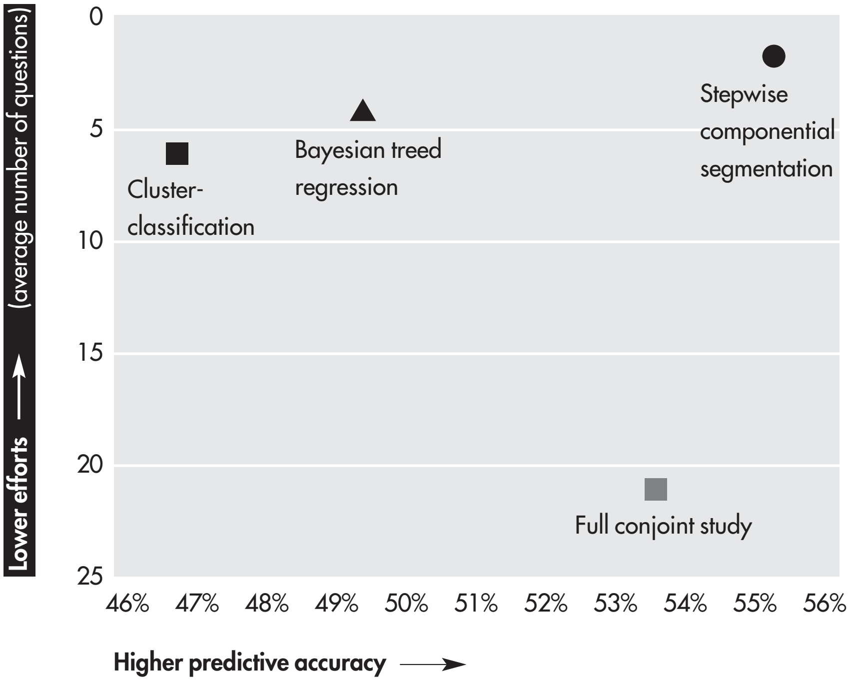
Question	**Full-profile conjoint**	**Cluster-classification**	**Bayesian treed regression**	**Stepwise componential regression**
Predictive accuracy	53.5%	46.9%	49.5%	55.1%
Average number of questions	21	6.2	4.2	2
Maximum number of questions	21	9	6	2
Incremental gain in predictive accuracy, per question (*)	1.4%	3.5%	5.8%	15.1%

(*) = (predictive accuracy – 25%) / number of questions. 25% is the predictive accuracy achieved by chance.

Figure 3

Predictive Accuracy vs. Number of Questions

Methods plotted on holdout sample predictive accuracy and respondents' efforts (number of questions). The stepwise compential segmentation, in the upper right corner, dominates the other methods on both dimensions.



Holdout sample test design

Preference partworths for the 100 individuals retained for holdout sample testing were calculated based on the analysis of the calibration sample.

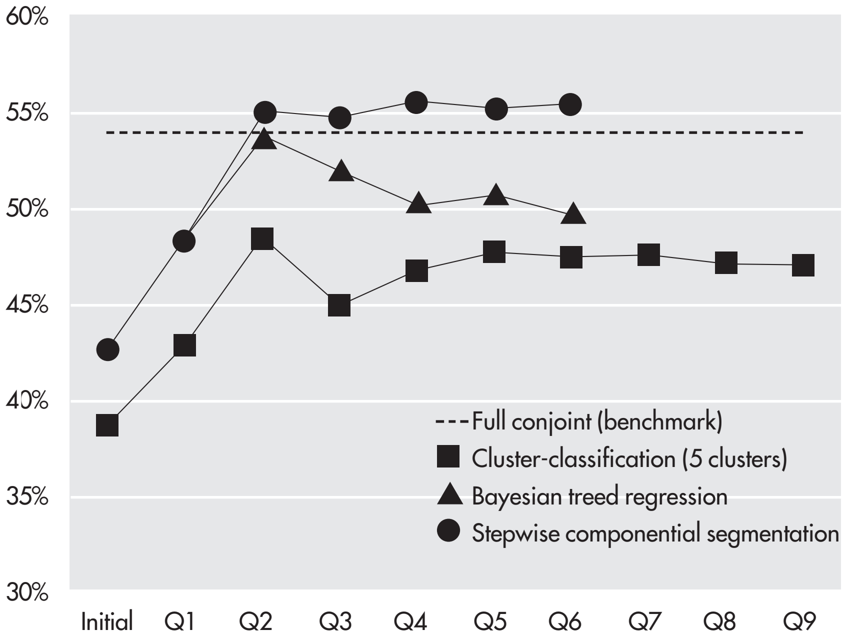
For the cluster-classification method, probabilities to belong to each of the five clusters were initially set equal to the proportion of each cluster in the in-sample population. Then, individuals in the testing set navigated the estimated tree based on their answers to the demographic and product usage questions, and probabilities of cluster membership were updated after each question that they answered. At each step, their estimated preference partworths were an appropriately weighted average of the preference partworths of the clusters to which they were assigned. Table 3 reports an example of how a specific individual answered the questionnaire, and how his probability to belong to the third cluster was updated based on his answers.

The same procedure was applied to the Bayesian treed regression method: individuals navigated the estimated tree, and their preference partworths were updated based on their answers. The Bayesian treed regression provides a different set of estimates for each end node of the tree. It is, therefore, straightforward to assign preference partworths to individuals based on the final node they occupy. For non-final

Figure 4

Holdout Sample Predictive Accuracy of the Three Competing Methods, after *N* Questions

Both stepwise componential segmentation and Bayesian treed regression achieve excellent predictive accuracy, but the latter eventually suffers from overfitting. Notice that the stepwise componential segmentation algorithm suggests to stop after two questions—predictive accuracy afterwards is only reported for comparison purpose.



nodes, however, no such estimates are offered (Bayesian treed regression optimizes the tree globally, that is, it fits data at the end nodes only. This explains why intermediary estimates are not reported). Intermediary partworths were computed as an appropriately weighted average of the partworths in the remaining "downward" nodes of the tree.

For the stepwise method, partworths were directly a function of the questions answered. Table 4 (page 10) indicates how preference estimates are updated after the first three questions (i.e., reports Ψ).

Results

In Table 5 (page 11) we report the holdout sample predictive accuracy of the three methods, namely the frequency with which each method correctly predicts the participant's top choice in the hold-out task that consisted of dividing 100 points among four alternatives. The holdout sample predictive accuracy of classic conjoint analysis estimates was 53.5% (after 21 questions), an improvement of 28.5% compared to chance.

The stepwise componential regression method, with a predictive accuracy of 54.6% achieved after only two questions, dominates the other methods, both in terms of efficiency and hold-out sample predictive accuracy (See Figure 3, page 11).

The closest contender is Bayesian treed regression, with an average predictive accuracy of 49.5% at its end nodes, reached after 4.2 questions on average. Despite conservative parameterization and one-third of the calibration set dedicated to overfitting diagnosis, the treed regression approach suffered from overfitting problems, as shown in Figure 4: its maximum predictive accuracy was achieved after only two questions, with 53.5%. But it failed to stop the splitting process, and its performance gradually deteriorated afterwards.

It is fruitful to compare this result to the hold-out sample accuracy of the stepwise componential segmentation method after an equal number of questions. Because preference partworths are computed on the entire dataset, the method does not overfit the data. On the other hand, the Bayesian treed regression's estimates are computed at the node level, i.e., on shrinking portions of the dataset, hence leading to overfitting.

Finally, the cluster-classification method, with a predictive accuracy of 46.9% and a much longer sequence of questions, fares far worse than the other two methods.

Conclusions

Online consumers tend to be impatient. It is essential, therefore, for merchant websites to offer relevant recommendations to their online visitors, quickly and with minimal consumer inputs. We proposed a framework by which companies can develop recommendation agents that are capable of providing high quality advice to first-time and impatient consumers. Specifically, we explored how traditional conjoint techniques (that have been used in the past to uncover product preferences of groups of customers to the benefit of marketing managers) could be leveraged to design recommendation agents without requiring the extensive and detailed inputs usually necessary for this kind of models. We tested alternative implementations of this approach and show that the stepwise componential regression method offers promise as a solution to this problem. ■

Acknowledgements

The authors gratefully acknowledge the Institute for the Study of Business Markets at the Pennsylvania State University for its financial support.

Note

[1] Notice that Likert-*x* question, for instance, can be operationalized in different ways. On one hand, the raw scores can be inserted into the descriptor matrix, thus resulting in a linear combination of x different modes. On the other hand, the population can be described in terms of binary splits, e.g., those who answered y ($< x$) or less are assigned a value of 0 while those who answered $y + 1$ ($\leq x$) or more are assigned a value of 1.This allows both linear and nonlinear relationships between consumers' characteristics and preference partworths to be used as predictors.

References

Ansari, Asim, Skander Essegaier, and Rajeev Kohli (2000), "Internet Recommendation Systems." *Journal of Marketing Research* 37 (August), 363–75.

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone (1984), *Classification and Regression Trees*. New York, N.Y.: Wadsworth.

Bucklin, Randolph E., and Sunil Gupta (1992), "Brand Choice, Purchase Incidence, and Segmentation: An Integrated Modeling Approach." *Journal of Marketing Research* 29 (May), 201–15.

Chipman, H., E. George, and R. McCulloch (2002), "Bayesian Treed Models." *Machine Learning* 48, 299–320.

De Angelis, Chris (2001), "Sampling for Web-Based Surveys." Presented to the Annual Marketing Research Conference, Atlanta, Ga.

DeSarbo, Wayne S., Alexandru M. Degeratu, Michael J. Ahearne, and Kim M. Saxton (2002), "Disaggregate Market Share Response Models." *International Journal of Research in Marketing* 19(3), 253–66.

DeSarbo, Wayne S., Richard L. Oliver, and Arvind Rangaswamy (1989), "A Simulated Annealing Methodology for Clusterwise Linear Regression." *Psychometrika* 54 (December), 707–36.

DeSarbo, Wayne S., Michel Wedel, and Marco Vriens (1992), "Latent Class Metric Conjoint Analysis." *Marketing Letters* 3 (July), 273–89.

Enos, Lori (2001), "Report: Five Keys for E-Tail Success." *E-Commerce Times*.

Green, Paul E. (1984), "Hybrid Models for Conjoint Analysis: An Expository Review." *Journal of Marketing Research* 21 (May), 155–69.

Green, Paul E., and Wayne S. DeSarbo (1979), "Componential Segmentation in the Analysis of Consumer Trade-Offs." *Journal of Marketing* 43(4), 83–91.

Green, Paul E., and Abba M. Krieger (1991), "Segmenting Markets with Conjoint Analysis." *Journal of Marketing* 55 (October), 20–31.

Green, Paul E., Abba M. Krieger, and Catherine M. Schaffer (1993), "An Empirical Test of Optimal Respondent Weighting in Conjoint Analysis." *Journal of the Academy of Marketing Science* 21 (Fall), 345–51.

Grenci, Richard T., and Peter A. Todd (2002), "Solutions-Driven Marketing." *Communications of the ACM* 45 (3), 65–71.

Gupta, Sachin, and Pradeep K. Chintagunta (1994), "On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models." *Journal of Marketing Research* 31 (February), 128–36.

Hagerty, Michael R. (1985), "Improving the Predictive Power of Conjoint Analysis: The Use of Factor Analysis and Cluster Analysis." *Journal of Marketing Research* 22 (May), 168–84.

Hoch, Stephen J., and David A. Schkade (1996), "A Psychological Approach to Decision Support Systems." *Management Scienc*e 42 (1), 51–64.

Huberty, C.J. (1994), *Applied Discriminant Analysis*. New York, N.Y.: Wiley and Sons.

__________ (1989), "Problems with Stepwise Methods: Better Alternatives." In *Advances in Social Science Methodology*,Vol. 1, ed. B. Thompson, 43–70. Greenwich, Conn.: JAI Press.

Huffman, C., and B. Kahn (1998), "Variety for Sale: Mass Customization or Mass Confusion?" *Journal of Retailing* 74 (4), 491–513.

Kamakura, Wagner A. (1988), "A Least Square Procedure for Benefit Segmentation with Conjoint Experiments." *Journal of Marketing Research* 25 (May), 157–67.

Kamakura, Wagner A., Michel Wedel, and Jagadish Agrawal (1994), "Concomitant Variable Latent Class Models for Conjoint Analysis." *International Journal of Research in Marketing* 11 (5), 451–64.

Ogawa, Kohsuke (1987), "An Approach to Simultaneous Estimation and Segmentation in Conjoint Analysis." *Marketing Science* 6 (1), 66–81.

O'Keefe, Robert M., and Tim McEachern (1998), "Web-Based Customer Decision Support Systems." *Communications of the ACM* 41 (3), 71–8.

Rencher, Alvin C. (1995), *Methods of Multivariate Analysis*. New York, N.Y.: John Wiley & Sons.

Resnick, Paul, and Hal R. Varian (1997), "Recommender Systems." *Communications of the ACM* 40 (3), 56–8.

Sawtooth (1991), "*ACA System*." Evanston, Ill.: Sawtooth Software.

Schafer, J. Ben, Joseph A. Konstan, and John Riedl (2001), "E-Commerce Recommendation Applications." *Data Mining and Knowledge Discovery* 5 (1/2), 115–53.

Toubia, Olivier, Duncan I. Simester, John R. Hauser, and Ely Dahan (2003), "Fast Polyhedral Adaptive Conjoint Estimation." *Marketing Science*, 22 (3), 273–303.

Vriens, M., Michel Wedel, and T. Wilms (1996), "Metric Conjoint Segmentation Methods: A Monte Carlo Comparison." *Journal of Marketing Research* 32, 73–85.

Wedel, Michel, and Wagner A. Kamakura (2000), *Market Segmentation, Conceptual and Methodological Foundations*, 2nd ed. Boston, Mass.: Kluwer Academic Publishers.

Wedel, Michel, and C. Kistmaker (1989), "Consumer Benefit Segmentation using Clusterwise Linear Regression." *International Journal of Research in Marketing* 6, 45–9.

Wedel, Michel, and Jan-Benedict E. M. Steenkamp (1991), "A Clusterwise Regression Method for Simultaneous Fuzzy Market Structuring and Benefit Segmentation." *Journal of Marketing Research* 28 (November), 385–96.

__________ (1989), "A Fuzzy Clusterwise Regression Approach to Benefit Segmentation." *International Journal of Research in Marketing* 6 (March), 45–59.

Wittink, D.R., M. Vriens, and V. Burhenne (1994), "Commercial Use of Conjoint Analysis in Europe: Results and Critical Reflections." *International Journal of Research in Marketing* 53(July), 41–52.

Report No. 05-103
"Offering Online Recommendations to Impatient First-time Customers with Conjoint-based Segmentation Trees" © 2005 Arnaud De Bruyn, John C. Liechty, Eelko K.R.E. Huizingh, and Gary L. Lilien; Report Summary © 2005 Marketing Science Institute