MANAGEMENT SCIENCE
Vol. 54, No. 1, January 2008, pp. 71–82
ISSN 0025-1909 | EISSN 1526-5501 | 08 | 5401 | 0071

informs

DOI 10.1287/mnsc.1070.0760

Predicting Product Purchase from Inferred Customer Similarity: An Autologistic Model Approach

Sangkil Moon
Department of Business Management, College of Management, North Carolina State University, Raleigh, North Carolina 27695, smoon2@ncsu.edu

Gary J. Russell
Department of Marketing, Tippie College of Business, University of Iowa, Iowa City, Iowa 52242, gary-j-russell@uiowa.edu

Product recommendation models are key tools in customer relationship management (CRM). This study develops a product recommendation model based on the principle that customer preference similarity stemming from prior purchase behavior is a key element in predicting current product purchase. The proposed recommendation model is dependent on two complementary methodologies: joint space mapping (placing customers and products on the same psychological map) and spatial choice modeling (allowing observed choices to be correlated across customers). Using a joint space map based on past purchase behavior, a predictive model is calibrated in which the probability of product purchase depends on the customer's relative distance to other customers on the map. An empirical study demonstrates that the proposed approach provides excellent forecasts relative to benchmark models for a customer database provided by an insurance firm.

Key words: marketing; recommendation system; spatial statistics; autologistic model
History: Accepted by Jagmohan S. Raju, marketing; received October 6, 2005. This paper was with the authors 9 months for 2 revisions.

Introduction

Product recommendation models are key tools in customer relationship management (CRM). An effective recommendation model contributes to the marketing goal of customer expansion by offering high-valued products to regular customers. This research develops an effective product recommendation system based on the principle that customer preference similarity stemming from prior purchase behavior is an important element in predicting product choice.

Our study brings together two complementary research methodologies: joint space mapping methodology (placing customers and products on the same psychometric map) and spatial choice modeling (allowing observed choices to be correlated across customers). The current research, however, differs from existing research in two respects. First, marketing science models based on psychometric maps typically examine distances between products and customers to infer the attractiveness of each product to each customer (Holbrook et al. 1982, Moore and Winer 1987, Gruca et al. 2002). In contrast, this research infers product preferences from intercustomer distances (Kapteyn et al. 1997, Yang and Allenby 2003). Because this approach only relies on customer positions on a psychometric map, it does not require that the researcher accurately locate a new product on an existing product map.

Second, contrary to previous work, this research uses a choice model adapted from the spatial statistics literature (Cressie 1993, Russell and Petersen 2000, Bronnenberg and Mahajan 2001, Bronnenberg and Sismeiro 2002). Simply put, spatial models assume that entities (such as customers) can be located in a space. Responses by entities are assumed to be correlated in such a manner that entities near one another in the space generate similar outcomes. Typically, the space in a spatial model is physical geography. Our research differs from the spatial statistics literature by analyzing a psychometric map rather than a physical map. Choice probabilities are predicted on the basis of the autologistic (AL) spatial model (Russell and Petersen 2000). This model, which takes a map as a key input, allows the researcher to construct a multivariate distribution of correlated binary (buy/no buy) response variables (Besag 1974, Cressie 1993). Because the AL model links these intercustomer correlations to distances on the psychometric map, we are able to infer product preferences based upon a customer's map location.

This paper is organized as follows. We begin by reviewing various product recommendation models in the management science and marketing science literatures. Next, we develop a product recommendation system based on the AL choice model. We show

how model calibration and forecasting can be carried out using a relatively simple maximum pseudo-likelihood estimation approach. By analyzing a customer database taken from an insurance firm, we demonstrate that the model has excellent predictive properties relative to a number of strong competitors. We conclude by discussing the contributions of this work and offering suggestions for future research.

Product Recommendation Models

Product recommendation models match customers to products using information that allows the firm to infer product preferences. Models have been proposed in both the management science and marketing science literatures. Below, we briefly review previous models, noting relationships to our proposed recommendation system.

Management Science Models

Many product recommendation models can be regarded as a type of collaborative filtering (Herlocker et al. 1999, Ariely et al. 2004). In collaborative filtering, choices of one customer are predicted using purchase information from other similar customers. For example, nearest neighbor algorithms, when used to implement collaborative filtering, are based on computing the similarity between customers based on their preference history using cluster analysis or the Pearson correlation coefficient. Predictions of the degree to which a customer will like an unknown product are computed by taking a weighted average of the opinions of a set of nearest neighbors for the target product.

In contrast, content filtering (Mooney and Roy 2000, Ariely et al. 2004) uses a given customer's product attribute ratings for other products to predict the customer's response for a product of interest. The key idea of content filtering is similar to the theory underlying multiattribute utility theory (Mazis et al. 1975) or conjoint analysis (Andrews et al. 2002): preferences for a product can be predicted by appropriately weighting the values of product attributes. Whereas collaborative filtering requires overall preference ratings for different products, content filtering requires customers' specific ratings for product attributes over different products. The lighter data demands of collaborative filtering makes the algorithm extremely popular for Internet retailers such as amazon.com and barnesandnoble.com.

Despite its high popularity in commercial applications, collaborative filtering (CF) is open to criticism (Ansari et al. 2000, Iacobucci et al. 2000, Melville et al. 2002). Many collaborative filtering techniques are based on customer preference ratings instead of actual purchases. Typically, subjects rate products very selectively. Because the subject-product ratings matrix can be very sparse, the likelihood of finding a set of subjects with significantly similar ratings is usually low. The recent interest in hybrid models (Balabanovic and Shoham 1997, Pazzani 1999), where collaborative filtering is combined with content filtering, is a reaction to this data sparsity issue.[1]

Marketing Science Models

These three types of algorithms—collaborative filtering, content filtering, and hybrid models—are ad hoc in the sense that they are not based on a statistical error theory. In contrast, recommendation models in the marketing science literature typically incorporate an error theory that recognizes that actual outcomes will fluctuate around predicted outcomes. Gershoff and West (1998) augment an individual-level multiattribute model with prediction residuals from other customers to yield more accurate forecasts for a given customer. This model uses features found in both content filtering and collaborative filtering: the basic utility specification is similar to a conjoint utility model, but information from other customers is used to predict the preferences of a given customer. Ansari et al. (2000) develop a hierarchical Bayes model in which the preferences of a given customer depend on both customer demographics and product attributes. Again, an individual is assumed to have a multiattribute utility model governed by customer-specific attribute weights, but pooling of information across customers permits more accurate forecasts. Hence, marketing science models use the same basic ideas as management science models, but express these ideas in a considerably different manner.[2]

Proposed Recommendation Model

Our proposed model can be regarded as a statistically-based collaborative filtering model in which the probability of purchase is the outcome measure. In common with collaborative filtering, our proposed model predicts choice by weighting the choices of customers by the degree of similarity between customers. However, because the model is based on an explicit statistical model of the correlations in choice outcomes among customers, we are able to use the underlying theory to define an appropriate measure

[1] Recently, Melville et al. (2002) proposed a two-step procedure to overcome data sparsity. First, they exploit content filtering to transform an original sparse matrix into a full matrix by predicting purchase/nonpurchase for nontested matrix cells. They then use collaborative filtering to predict holdout cases using the less sparse transformed matrix.

[2] Ansari and Mela (2003) have developed a statistical model of collaborative filtering using a Bayesian semiparametric probit analysis with a Dirichlet process prior. However, the purpose of this model is to increase website traffic as opposed to making product recommendations.

of customer similarity. Moreover, because the procedure incorporates a psychometric map based on past purchases, we are able to infer customer preference for a target product without reliance on a specific multiattribute utility model or the need to define product attributes. We show subsequently that the use of past purchase information in a spatial choice framework leads to an effective forecasting system.

Autologistic (AL) Recommendation Model

A product recommendation model is a forecasting system that uses information in a customer database to forecast the probability that a customer will buy a product. It is assumed that a database marketer would like to cross-sell a particular target product to regular customers who have not yet been asked to buy the product (Kamakura et al. 1991, 2003; Bodapati 2008). The focus here is constructing a model that allows the marketer to identify customers in the database who would be good prospects for a target product solicitation.

Conceptual Overview

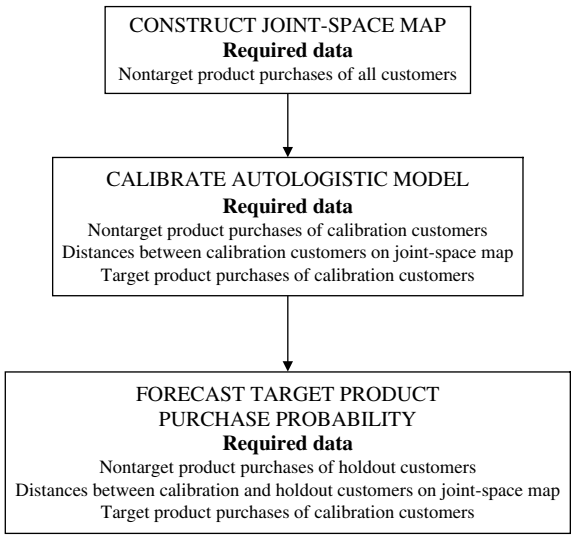
We assume that we have access to a customer database consisting of two groups: customers who have had an opportunity to buy the target product (calibration customers), and customers who are currently unaware of the product (holdout customers). Information is available on purchases of nontarget products for all customers. The goal is to use the customer background information (for all customers) and actual purchases of the target product (for calibration customers) to predict the probability that each of the holdout customers will buy the target product.

As shown in Figure 1, the procedure involves three steps. In the first step, we construct a joint-space map using the nontarget product purchase histories of all customers in the database. This map contains an ideal point for each customer and a location for each nontarget product. In the second step, we build a predictive model based on the joint-space map and the actual purchases of the calibration customers. As will be seen subsequently, this model links the estimated purchase probability of the focal customer to the actual purchases of other customers using a measure of customer similarity derived from the joint-space map. In the final step, we use the predictive model to forecast the probability that each holdout customer will buy the target product.

Joint-Space Mapping

The proposed model is based on a measure of customer similarity (CS) derived from a joint-space map in which customer ideal points and products are simultaneously represented. A pick-any dual scaling algorithm is used to generate a joint space based on customer purchase records (Levine 1979, Holbrook et al. 1982). We construct a customer by product matrix in which binary variables (buy/no buy) indicate whether or not a customer has purchased a particular product in the past. Because we do not have sufficient information to locate the target product, we exclude the target product in constructing the matrix. An eignenvector analysis of the customer by product matrix then creates a map in which customers are located in the center of the products that they already purchased. Details on the algorithm can be found in Appendix A.

Figure 1 Conceptual Overview of the Model



Conceptually, the pick-any procedure estimates the map position of customers on the basis of similarity in product preferences. We define the preference similarity of two customers based on the proximity of their ideal points on the map. That is, for any two customers (a and b), we can compute the distance between customer ideal points as the Euclidian metric

$$CD_{ab} = \sqrt{\sum_r (a_r - b_r)^2}, \tag{1}$$

where the summation runs over each dimension r of the constructed joint space. Assuming that the number of customers exceeds the number of products, the maximum dimension of a pick-any map is $K - 1$, where K is the number of products included in the customer by product matrix (Moon 2003).

It should be noted that the map distances are treated as input information for our predictive model. The goal of this research is not to calibrate a product map while simultaneously forecasting choice. Rather, our goal is to use an externally generated psycho-

metric map as a way of measuring similarity among customers. Viewed broadly, a researcher implementing our approach has complete freedom to develop alternative measures of similarity using other types of information (such as psychographic or lifestyle variables).

Autologistic (AL) Model

The second step of our procedure is to develop a model that uses patterns in the calibration customer group to infer the target product choice probabilities for the holdout customers. In this step, we make use of a spatial statistical model in which the probability of purchase of one customer is linked to the actual choices of other customers. The key assumption of our model is that customers who are near one another on the joint-space map have similar preference structures.

The model used here is a modified version of the AL regression model found in the spatial statistics literature (Besag 1974, Cressie 1993, Russell and Petersen 2000). Let Z_c be a binary variable indicating whether or not customer c has purchased the target product. The AL model assumes that the probability that customer c purchases the target product is given by the conditional logit model

$$\Pr(Z_c = 1 \mid Z_{c'} \text{ for } c' \neq c) = \left[1 + \exp\left[-\left\{\alpha + \sum_{c' \neq c} \theta_{cc'} Z_{c'}\right\}\right]\right]^{-1}, \quad (2)$$

where the $\theta_{cc'}$ are parameters that measure the similarity between the focal customer c and another customer c'. The conditional logit model in Equation (2) implies that the outcome variables Z_c are correlated across customers. Conceptually, the purchase probability of focal customer c is determined by other customers' past purchase behavior. If there are many target product buyers near the focal customer c on the map, our model predicts that the focal buyer is likely to buy the focal product as well. Formally, Equation (2) states that we can predict purchase probability if we know the relative similarity of customer c to other customers ($\theta_{cc'}$) and whether the other customers have purchased the target product ($Z_{c'}$).

In statistical terminology, Equation (2) is called a *full conditional distribution*: it specifies the probability distribution of Z_c (for customer c) given the known values $Z_{c'}$ (for all other customers c'). Because these full conditional distributions (one for each customer in the database) are mutually dependent, we cannot analyze this model further unless we understand the joint distribution of all the Zs. Put another way, the AL expression summarized by Equation (2) is a way of specifying a simultaneous equation system that predicts the choice outcomes for all customers in the database.

The form of the joint distribution of all the Zs is not evident from simple inspection of Equation (2). However, by applying a key theorem due to Besag (1974) and assuming that all intercustomer parameters are symmetric ($\theta_{cc'} = \theta_{c'c}$), it can be shown that the joint distribution of the Z_c ($c = 1, 2, \ldots, N$) is given by the multivariate logistic distribution of Cox (1972). Detailed information on the multivariate logistic distribution and its relationship to the AL expressions in Equation (2) can be found in Appendix B. (Additional discussion about the AL model and the multivariate logistic distribution is presented by Cressie 1993 and by Russell and Petersen 2000.) It should be noted that the symmetry of the $\theta_{cc'}$ coefficients is a key element in the model specification. If the $\theta_{cc'}$ are not symmetrical, the methodology fails because no joint distribution of the Z_c variables exists (Besag 1974).

Defining Customer Similarity

In the multivariate logistic distribution, the $\theta_{cc'}$ parameters are functions of the correlations among the Z_c variables. In general, due to the fact that the correlations of the multivariate logistic distribution are unrestricted, the signs of the $\theta_{cc'}$ parameters can be positive, negative, or zero. However, in our application, we assume spatial continuity of preferences on the pick-any map (Haining 1990): customers near one another on the map are assumed to have similar preferences. This assumption is reasonable in our model because the pick-any mapping procedure places customers with similar purchase histories near each other on the map (see Appendix A). For this reason, we constrain the values of $\theta_{cc'}$ to be nonnegative and interpret the $\theta_{cc'}$ parameters as symmetrical, ratio-scaled measures of customer similarity.

Because the number of distinct $\theta_{cc'}$ parameters equals $N(N-1)/2$, where N is the number of calibration customers in the database, it is infeasible to calibrate the model without further restrictions. We solve this problem by linking $\theta_{cc'}$ to the proximity between customers on the joint-space map. Specifically, we assume that

$$\theta_{cc'} = \beta \exp(-\lambda CD_{cc'}), \quad (3)$$

where $CD_{cc'}$ is the distance defined in Equation (1), and β and λ are parameters to be estimated. Here, the λ parameter ($\lambda > 0$) scales the map distances and determines the strength of dependence across customers. The negative exponential function is based on the continuity principle: nearby customers tend to have similar preferences. Correlation in choices (as defined by the multivariate logistic distribution of the

binary purchase variables) declines as the distance between customers on the map increases.[3]

Using Equation (3), the specification of the AL model can be written in the compact form

$$\Pr(Z_c = 1 \mid Z_{c'} \text{ for } c' \neq c) = [1+\exp[-\{\alpha+\beta CS_c\}]]^{-1}, \quad (4)$$

where the implied customer similarity variable

$$CS_c = \sum_{c' \neq c} \exp(-\lambda CD_{cc'}) Z_{c'} \quad (5)$$

depends on both the intercustomer map distances and whether or not other customers c' have purchased the product. Two features of the model should be emphasized at this point. First, the number of model parameters *does not* increase with the number of customers in the database. Conditional on the pick-any map, the customer similarity variables in (5) are known variables that can be computed from the data set. Second, working with the full conditional distributions as specified in Equations (4) and (5) is equivalent to specifying the structure of a multivariate distribution of binary purchase variables across the calibration customer population. However, the intuition of the model—linking purchase probability to the purchase behavior of similar customers—is best understood by the conditional model specification given above.

Calibrating the Autologistic (AL) Model

Direct use of the multivariate logistic distribution for parameter estimation is mathematically intractable. Accordingly, we follow the common practice in the spatial statistics literature of using the full conditional distributions in Equations (4) and (5) to calibrate the model parameters. Note that the logit expression in Equation (4) is not independent across customers because each customer's purchase variable Z_c is related to the purchase variables of all the other calibration customers. Consequently, from the standpoint of statistical theory, we should not multiply together the N binary logit likelihoods generated by Equation (4) to compute the overall likelihood function. Nevertheless, maximizing this function leads to consistent estimates of model parameters under certain conditions. This procedure, known as maximum pseudo-likelihood estimation (MPLE), has been used to calibrate model parameters in many studies involving spatial data (Cressie 1993).

Proof of consistency of the MPLE procedure has been established for some types of spatial models (Strauss and Ikeda 1990, Arnold and Strauss 1991, Aerts and Claeskens 1999, Arnold et al. 2001, Molenberghs and Verbeke 2005). However, due to the unique properties of the pick-any scaling methodology, we are unable to offer a formal argument for consistency in our application. Instead, we regard MPLE as a reasonable calibration algorithm for the AL model, and focus attention on the ability of the AL model to forecast choice behavior. We note that data mining procedures (such as neural networks) that are calibrated using ad hoc fitting algorithms often yield useful forecasts (Hastie et al. 2001).

Forecasting Holdout Customer Response

After the model is calibrated, forecasting the probability of purchase for the target product in the holdout customer data set is straightforward. We first compute the CS measures for each holdout customer as in Equation (5). In this case, however, the summation in each CS measure runs over all the customers in the calibration data set. (Clearly, we do not know whether or not the other holdout customers will purchase the target product.) Once all the CS measures are defined, the AL model of Equation (4) is used to compute choice probabilities for each customer in the holdout sample.[4]

Application

In this section, we use the AL model to predict the reactions of customers to a solicitation to buy an insurance policy. Because the actual choices of all customers are known in this application, we are able to use these data to study the effectiveness of the proposed model in predicting choice. Results indicate excellent performance for the AL model.

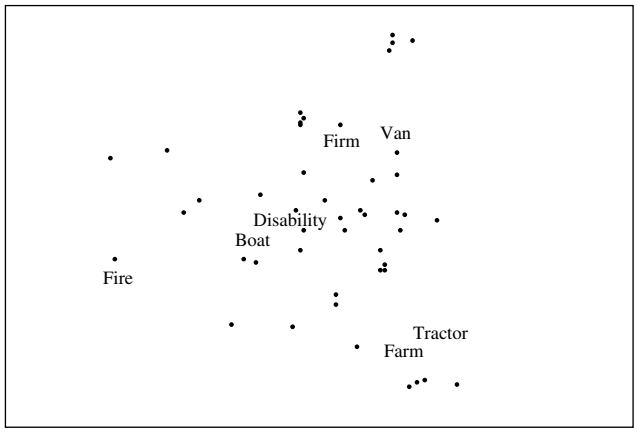
Data Description

The data used in this study were taken from the database of a European insurance firm. The database provides variables related to sociodemographics and insurance policy ownership. Sociodemographic variables are based on the postal code of the residence of the customer. That is, all the people in the same postal code area have the same sociodemographic characteristics. One insurance policy (private third-party insurance) was selected as the target product for recommendation in this analysis. This type of insurance is similar to personal property insurance (for the contents of homes) in the United States. Generally, there were low correlations between the product ownership variables and the demographics variables. We selected a random sample of 4,000 customers for the analysis.

[3] It should be noted that $\theta_{cc'}$ is not the correlation between the purchase outcomes of customers c and c'. However, the correlation matrix for all purchase outcomes is determined by the set of all $\theta_{cc'}$ values.

[4] This approach implicitly assumes that we have a sample of $(N+1)$ customers: N calibration customers and one holdout customer. The forecast for the holdout customer is the full conditional distribution of this customer's response variable conditional on the responses of the N calibration customers. We then repeat this process for each holdout customer in turn.

Figure 2 First Two Dimensions of the Pick-Any Map



Notes. The AL model is based on a six-dimensional pick-any map. Only the first two dimensions are shown here. Because the target product (private third-party insurance) is not used to create the map, it does not appear in this figure. Customer ideal points are denoted by dots. Although product points assist in the interpretation of the map, they are not used in choice prediction. The AL model only uses distances between customer ideal points.

Constructing the Joint-Space Map

We used seven insurance products (out of a possible set of 21 nontarget products excluding the target product) to construct the joint-space map. These seven products were selected on the basis of a stepwise logit regression for the calibration customers, in which all the nontarget insurance policy ownership variables were used to predict the ownership of the target insurance policy. We retained only those insurance ownership variables that were statistically significant at the 0.05 level. This screening procedure is essential to ensure that the map represents a product cluster that is useful for predicting the target product (Solomon and Buchanan 1991).

The 4,000-by-7 (customer-by-product) matrix of binary variables (buy/not buy) denoting customer product ownerships was analyzed using Levine's (1979) pick-any scaling algorithm (see Appendix A).[5] In Figure 2, we display a plot of the first two dimensions of the pick-any map. The actual map used in model development contained six dimensions.[6]

Note that the map contains locations for both insurance policies and customer ideal points. The target product (private third-party insurance) is missing from Figure 2 because we do not use information on the target product in creating the pick-any map. As discussed earlier, only the intercustomer distances play a role in our recommendation model. Product positions are generated by the pick-any mapping algorithm, but are ignored by the AL model. However, product positions serve to emphasize the fact that customer ideal points are located in the region of the space near insurance policies that have been previously purchased. Accordingly, customer similarity in the AL model is defined with respect to previous purchase behavior.

The map in Figure 2 is interesting because it provides some insights into the purchase behavior of the households in our study. Roughly speaking, the map has three regions: business insurance (top right), agricultural insurance (bottom right), and personal property insurance (middle left). We reran the pick-any analysis including the target category to determine where it would be located on a pick-any map. This analysis (not shown) suggests that private third-party insurance would have a product position very close to fire insurance. Managerially, the pick-any map clearly shows how customer characteristics impact the type of insurance policy purchased. The AL model procedure exploits this revealed segmentation in generating predictions of buying behavior.

AL Model Specifications

In our empirical work, we use the basic AL model (Equations (4) and (5)) in which nontarget insurance policy purchase variables impact purchase probability through the customer similarity (CS_c) variable. Because this basic model predicts target product purchase probability solely on the basis of prior purchase (denoted P) information, we denote the model using the notation AL-P. It should be noted that sociodemographic variables are not used in creating the map shown in Figure 2. To study the role of demographics in predicting choice, we also considered two variants of the AL model. The AL-D model uses a CS index based on demographics alone. In calculating customer demographic similarity, the correlation between the demographic profiles of two customers replaced the CS_c term in Equation (5). In contrast, the AL-PD model combines both past purchases and demographics in the hybrid structure

$$\Pr(Z_c = 1 \mid Z_{c'} \text{ for } c' \neq c) = \left[1 + \exp\left[-\left\{\alpha + \beta_1 CS_c^{(P)} + \beta_2 CS_c^{(D)}\right\}\right]\right]^{-1}, \quad (6)$$

[5] The most commonly used pick-any scaling algorithm (described in Appendix A) requires the researcher to analyze a data matrix whose dimension is $(K + N)$ by $(K + N)$, where K is the number of products and N is the number of customers. Considering the large number of customers in a typical customer database, this method limits the scalability of the procedure. To overcome this problem, we adopted a reduced matrix approach discussed by Levine (1979, p. 88) and Holbrook et al. (1982, p. 100). The reduced matrix approach analyzes a K by K matrix instead of the usual $(K+N)$ by $(K+N)$ matrix. In our application, $K = 7$ and $N = 4{,}000$.

[6] As noted earlier, the maximum number of dimensions of a pick-any map is $K - 1$, where K is the number of products. In this research, we decided to use all $K - 1$ dimensions in computing customer distances. There exists some evidence that choosing a lower-dimensional solution can increase the forecasting ability of a predictive model based on a pick-any map (Kwak 2004). Accordingly, use of all $K - 1$ dimensions can be viewed as a conservative test of the AL recommendation model.

where $CS_c^{(P)}$ is the customer similarity measure based on purchases and $CS_c^{(D)}$ is the customer similarity measure based on demographics. In this model, each source of CS is weighted differentially by the size of the β parameters. Clearly, the AL-PD generalizes the definition of customer similarity to encompass multiple sources of information. These generalizations of the basic AL model serve to emphasize the fact that the researcher has considerable latitude in selecting a suitable model specification for a particular application.

Prior to model calibration, we standardized the CS_c variables using the mean and standard deviation of these variables for customers in the calibration data set. This procedure controls the magnitude of the AL model parameters, but has no impact on forecast performance.

Benchmark Model 1: Principal Components Logit (PCL)

Four benchmark models are used to evaluate the performance of the proposed AL model. The first benchmark model, called the principal components logit (PCL) model, is a simple logit regression in which principal component variables are used to forecast the probability of target product choice. To construct this model, we first analyze all socio-demographics and nontarget insurance policy ownership variables using principal components. In doing this, we retain 22 principal components, accounting for 70% of the total variation of the original 62 raw independent variables (21 existing products variables and 41 demographics variables). Using these 22 principal components variables, we estimate the coefficients of a logit regression model using data from the calibration customers only. To forecast the choice probability of holdout customers, we inserted the value of each holdout customer's principal component variables (factor scores) into this logit model specification.

Benchmark Model 2: Stepwise Logit (SWL)

The second benchmark model is a stepwise logit (SWL) regression model that has prior purchase records as independent variables. In the model, we select only existing product variables statistically significant at the 0.05 level out of the original 21 existing products. The model is expressed as

$$\Pr(Z_c = 1) = \left[1 + \exp\left[-\left\{\alpha + \sum_k \gamma_k X_{kc}\right\}\right]\right]^{-1}, \qquad (7)$$

where X_{kc} are binary variables indicating whether customer c previously purchased the nontarget insurance policy k.

As with the AL model, three different SWL variants were examined. SWL-P is the SWL model in which the customer's past purchases were used as independent variables (as in Equation (7)). SWL-D is the SWL model in which the customer's demographics were used as independent variables. SWL-PD is a hybrid procedure in which the model makes use of the full database containing both past purchase indicators and demographics.

Benchmark Model 3: Collaborative Filtering (CF)

The third benchmark model, collaborative filtering (CF), is drawn from the management science literature (Schafer et al. 2001). This benchmark model is especially important because it is similar to models adopted by many Internet retailers. In this model, a Pearson correlation coefficient is used to compute customer similarity based on available customer information. Specifically, the model takes the form

$$[\text{Forecast } Z_c] = b + \sum_{c' \neq c} r_{cc'}(Z_{c'} - b)/K_c, \qquad (8)$$

where $r_{cc'}$ is the correlation between customers c and c', $K_c = \sum_{c' \neq c} r_{cc'}$, and b is the base rate of the response. We define the base rate to be the overall proportion of calibration customers who buy the target product.

Two aspects of the CF model should be noted here. First, the CF model is not estimated by a statistical procedure. More specifically, it has neither an error term nor a parameter to estimate. Accordingly, once correlations are computed, Equation (8) is used to compute the expected value of Z_c. Because Z_c is a binary (0–1) variable in this study, we regard [Forecast Z_c] as an estimate of the probability that customer c will purchase the target product. Second, the summation over customers c' is interpreted to mean summation over a small neighborhood of customers with high correlations relative to customer c. For this study, we define the neighborhood as the top 7.5% closest customers (as measured by correlation) to the target customer. This definition, developed empirically, maximizes the performance of the CF procedure in our data set.

As with the AL model, three different CF variants were examined. CF-P is a collaborative filtering model in which the customer's vector of past purchases is used to compute the correlation coefficients among customers. Note that past purchases include binary variables indicating purchase status of the seven nontarget product categories. CF-D is a collaborative filtering model in which the customer's vector of demographic variables is used to compute correlation coefficients. CF-PD is a hybrid procedure in which correlations are computed using a vector containing both past purchase indicators and demographics.

Benchmark Model 4: Neural Network (NN)

Neural-network (NN) models are known for superior predictive performance on problems involving many

variables and extreme curvature in the response function (Soucek 1991, Hastie et al. 2001, Kim et al. 2005). For the purposes of comparison with the AL-P model, we calibrated a NN-P model using the seven nontarget products used for the pick-any map. Due to the relatively small number of inputs, we used an NN architecture consisting of one hidden layer with four nodes. The functions linking the input nodes and the hidden layer nodes were assumed to be in linear form (the sum of weight times input value). The functions linking the hidden layer nodes and the output nodes were assumed to be in logistic form. As in the AL model and CF model groups, we also estimated two other NN variants. The NN-D used only demographic variables as inputs. The NN-PD used both demographic and past purchase variables as inputs. The NN-D model had one hidden layer with four nodes, while the NN-PD model had one hidden layer with eight nodes.

Analysis of Model Performance

Following standard practice in the data mining literature, model performance was assessed using a 10-fold cross-validation procedure (Hastie et al. 2001). As noted earlier, our data set includes 4,000 customers. This data set was randomly divided into 10 sections, each containing 400 customers. We then estimated each model 10 times, each time using nine sections for calibration (3,600 customers) and one section for forecasting (400 customers). Note that this procedure uses each customer for forecasting only once. The mean performance across the 10 forecast groups is the basis for our conclusions.

To give some indication of the properties of the MPLE model calibration procedure applied to the AL model, we provide the average AL parameters obtained in the cross-validation analysis in Table 1. We note that all coefficients on CS for both purchases and demographics in the AL model variants have the expected positive signs. The mapping scale parameter λ also has the theoretically correct positive sign. Analysis of the map distances (not shown) indicates that forecasts of choice probability are heavily weighted toward those customers with very similar patterns of past purchases. We provide the standard deviation for each parameter to provide an indication of the stability in the parameter value across the different calibration data sets. Because we regard MPLE only as a reasonable fitting algorithm, we do not compute standard errors of model parameters.

Table 1 Parameter Estimates of AL Models

	AL-P	AL-D	AL-PD
Intercept	−0.4203 (0.0464)	−0.2617 (0.0452)	−0.4198 (0.0409)
Customer preference similarity	1.3726 (0.0190)	—	1.3743 (0.0177)
Map scale parameter (λ)	0.7870 (0.0670)	—	0.7200 (0.0589)
Customer demographic similarity	—	0.0605 (0.0075)	0.0612 (0.0079)

Notes. This table lists the mean and standard deviation (in parentheses) of the parameter estimates of the AL models across the 10 calibration samples of the cross-validation analysis. Model codes are in the form of AL-YY, where YY is the data source. Data codes are P, product purchases; D, demographics; PD, product purchases and demographics.

Model Performance Measures

To test model performance, we use three different measures: (1) mean absolute deviation (MAD), (2) hit rate (HR), and (3) the Gini coefficient. The MAD statistic measures how accurately the estimated purchase probabilities match actual purchase behavior. We define MAD as

$$\text{MAD} = \sum_c |\Pr_c - Z_c|/N, \tag{9}$$

where $\Pr_c$ indicates the estimated purchase probability for customer c, Z_c the binary target product purchase variable (1 = purchase; 0 = no purchase), and N the number of customers in the sample. In words, MAD is the average of the difference between the estimated purchase probability and the actual target product purchase variable over all the customers in the sample. We also examine HR. In this study, HR is defined as the ratio NF/NB, where NF is the number of buyers in the sample whose estimated purchase probability is larger than 0.50, and NB is the total number of buyers in the sample. This index measures the ability of the model to correctly forecast purchase behavior among the group of consumers who ultimately will buy the product.

Adopting industry practice, we also examined the Gini coefficient in evaluating the predictive performance of the various models. The Gini coefficient is determined by the area between each model's cumulative lift curve and the random assignment (RA) lift curve.[7] Because the best-possible lift curve varies as a function of the overall purchase rate among customers in the holdout sample, the Gini coefficients in our study have been normalized so that a value of 1.0 corresponds to the best-possible lift curve. Higher

[7] Lift is defined as the percentage of all buyers in the holdout data set who fall into the group of customers selected for purchase solicitation. For each model, purchase probabilities of the holdout customers are forecasted. Holdout customers are then ordered in terms of forecasted choice probability, and the top x% of the holdout customers are selected, where x% corresponds to the selection rate. For example, we can vary x from 0 to 100 at intervals of 10 (i.e., 0%, 10%, 20%, etc.).

Table 2 Measures of Holdout Sample Model Performance

Information set	MAD	HR	Gini
Purchases only			
AL-P	**0.319**	0.862	**0.670**
NN-P	0.326	0.871	0.588**
SWL-P	0.331**	**0.874****	0.652
CF-P	0.344**	0.847*	0.606
Demographics only			
AL-D	0.482	0.372	0.086
NN-D	**0.469***	**0.499****	**0.104**
SWL-D	0.478	0.188**	0.064
CF-D	0.470*	0.245**	0.058*
Purchases and demographics			
AL-PD	**0.310**	0.868	**0.703**
NN-PD	0.328**	0.837**	0.566**
SWL-PD	0.331**	**0.874**	0.621*
CF-PD	0.429**	0.502**	0.486**
PCL-PD	0.406**	0.593**	0.467**

Notes. Model names: Model codes are in the form of XX-YY, where XX is the model name and YY is the data source. Model codes are AL, autologistic; NN, neural network; SWL, stepwise logit; CF, collaborative filtering; and PCL, principal component logit. Data codes are P, product purchases; D, demographics; PD, product purchases and demographics. Model performance: This table displays the mean performance of each model over the 10 holdout samples of the cross-validation analysis (MAD, mean absolute deviation; HR, hit rate; Gini, Gini coefficient). Entries in bold indicate the best performance of any model within a particular information set. Asterisks indicate the results of a paired *t*-test comparing the AL model of a particular information set to alternative models within the same information set.

*Statistical significance at the 0.10 level.

**Statistical significance at the 0.05 level.

values imply a better overall model. Intuitively, the Gini coefficient is a measure of the ability of a model to correctly rank order customers in terms of their likelihood of buying the target product.

Model Comparison

A summary of the cross-validation analysis of various models is displayed in Table 2. For each model, we show the average model performance statistic across the 10 forecast samples. We also provide the results of a paired *t*-test in which the AL model is compared to all other models using the same information source.[8] For example, in the "Purchases only" section of the table, the AL model based on only purchase information is compared to other models that also use only the purchase information. We also indicate, using boldface, the model that has the best performance on each measure in terms of the mean. Based on our earlier discussion, a good model has small values of MAD and large values of both HR and Gini coefficient.

[8] The paired *t*-test controls for correlations across methods within the same forecast data set, but does not address potential correlations across forecast data sets. We leave the development of a stronger test methodology to future research.

Two conclusions emerge from the summary of model performance displayed in Table 2. First, the forecasting ability of past purchases far exceeds the forecasting ability of demographics alone. Note that a Gini coefficient equal to zero indicates a random sorting of households in terms of purchase probability. Although there are differences in procedures (generally favoring the neural network model), the extremely low Gini coefficients make it clear that no model based solely on demographics forecasts purchase probability particularly well. We note that the superiority of purchase history over demographics is commonly found in marketing science studies (see, e.g., Rossi et al. 1996). However, the results here must be viewed cautiously. Recall that we only know the average demographics of the postal code of each customer. This mismatch between the customer's actual demographics and the reported demographics in the database probably plays a key role in the poor performance of demographic information.

Second, the AL model performs well when purchase information is part of the information set. In terms of raw means, the AL model has the smallest MAD and largest Gini coefficient. These differences are statistically different (better) than all other models when both purchase history and demographics are used to model choice behavior. In terms of HR, the best performance in Table 2 was generated by the stepwise logit model (SWL), but the HR measures for the AL models are quite reasonable. In fact, the AL model HR mean is not statistically different than the SWL model HR mean when both purchase history and demographics are used.

When interpreting the model fit summary in Table 2, it is useful to keep in mind that the three measures represent different aspects of model performance. From the standpoint of the manager, the two most important measures are clearly MAD and the Gini coefficient. Most marketing policies in a database marketing context are based on either an expected profitability calculation (requiring an accurate estimate of purchase probability) or a rank order of purchase likelihood for all customers in a database. Because MAD assesses the accuracy of the estimated purchase probability and the Gini coefficient assesses accuracy of the rank order, the AL approach appears to be well suited to database marketing applications in which past buying behavior is used to forecast future buying behavior.

Summary

This application provides substantial evidence in support of the spatial model approach in developing a product recommendation model. In theory, the space used in spatial modeling is a replacement for missing variables that actually drive response behavior (Cressie 1993, Bronnenberg and Mahajan 2001,

Bronnenberg and Sismeiro 2002). The joint-space map obtained from purchase information represents clusters of (unknown) variables that determine purchase behavior (e.g., word-of-mouth, network externalities, lifestyle). By relying upon this map, the AL spatial model offers better predictive performance using the same customer background information as other procedures.

Conclusion

This research develops a product recommendation system based on tools from the spatial statistics literature. In our proposed AL procedure, customers are assumed to be located on a joint-space (pick-any scaling) map in which individuals who are near one another share similar product preferences. In contrast to multidimensional scaling (which is used to determine how competing brands are perceived), pick-any scaling has been used to understand the relationship between brand positioning and customer preferences (Holbrook et al. 1982, Gruca et al. 2002). This study exploits the properties of pick-any scaling to derive a measure of customer similarity based on past purchase behavior. By merging the AL spatial choice model with a pick-any map, we are able to construct a new method of forecasting product choices for customers in a firm's database. We show empirically that the AL model provides excellent forecasts relative to benchmark procedures.

Managerial Implications

The AL model assists managers in two ways. First, the approach promises to be very effective in capturing the effects of variables that drive choice behavior, but are not explicitly included in the customer records available to the researcher. For example, variables such as lifestyle, psychographics, financial resources, and product features often determine choice behavior, but are typically absent from a firm's database. As long as a map can be created in which locational proximity is related to product preference, the spatial methodology allows the analyst to construct a model that implicitly contains more information about customer behavior than is apparent from the available data. In a certain sense, the spatial choice methodology can be interpreted as a representation of unobserved heterogeneity in customer preferences. However, instead of using statistical distributions to represent heterogeneity, we instead rely on a weighting of the choice behavior of other customers. The proposed model is flexible in the sense that it can integrate various sources of customer similarity information and differentially weight each source of information in explaining choice behavior.

Second, because the mapping procedure produces a representation of the relationships between customer ideal points and product purchases, managers can gain insights into the overall structure of the market. In Figure 2, we observe that insurance products cluster due to customer needs (insurance for personal property, farm operations, and business operations). Intuitively, the AL model can be regarded as a method of segmenting customers in terms of product preferences and then using this segmentation information to make forecasts of the purchase probability of additional products. These insights are more difficult to obtain using data mining tools such as neural networks or collaborative filtering.

Model Extensions

The AL model can be extended in a variety of ways. The most obvious extension would be a recommendation system in which forecasts are prepared simultaneously for multiple target products. In this setting, the researcher is assumed to be interested in predicting the likely purchase probabilities of a portfolio of products. This entails considering a model of binary variables Z_{cp} in which the subscript c denotes a customer and the subscript p denotes a target product. Embedding this structure in the multivariate logistic model implies that the researcher must consider the correlation linking multiple customers and multiple products. In practical terms, this model can improve forecasting by using the likelihood of purchasing in one category to impact the likelihood of purchasing in another category. This approach could prove especially useful if the products in the portfolio of purchases tend to be strong complements.

A further generalization to incorporate temporal dynamics is desirable. This would entail considering a model of binary variables Z_{cpt} in which the subscript t denotes time. Adding a temporal dimension (e.g., customers' dynamic learning, variety seeking) to the model involves a specification of temporal dependence in addition to correlations between customers and between target products in terms of customer purchase behavior. This generalization will be especially relevant to diffusion models, where we need to predict product adoption time in addition to adoption itself (Haining 1990, Banerjee et al. 2004).

Clearly, there will be little or no information on entirely new target products with regard to customers' responses. Accordingly, the proposed AL procedure is not as suitable for new target products as for existing target products. Analyzing similar products or similar markets may help predict who will purchase a new target product in the new market. To effectively apply the AL model to such new target products, marketers could select a small number of customers (calibration sample) from the database and expose them to the target product by a test promotion. Using purchase response from this test, the AL method could then be applied.

Acknowledgments

The authors thank participants of the Spatial Models in Marketing session at the 2004 Invitational Choice Symposium, Wagner Kamakura (Duke University), and Nick Street (University of Iowa) for providing insightful comments about this research. The authors also thank the associate editor and three anonymous reviewers for comments that materially improved the quality of this paper. Financial support from the Marketing Science Institute (MSI), the North Carolina State University Edwin Gill Research Grant, and the Center for Customer Relationship Management of Duke University is gratefully acknowledged. The first author won the 2002 MSI Clayton Doctoral Dissertation Proposal Competition Award for the work reported in this paper.

Appendix A. Generating a Joint-Space Map by Pick-Any Scaling

Pick-any scaling (Levine 1979) creates a map with two properties: customers are located in the *center* of the products that they select, and products are located in the *center* of customers who select them. (The meaning of *center* is defined below.) The mapping algorithm can be formulated as an eignenvalue problem as follows. Let N be the number of customers and K be the number of products. Define E as a symmetric matrix of dimension $N+K$ with the following form:

$$E = \begin{bmatrix} \phi_N & B \\ B' & \phi_K \end{bmatrix}, \tag{A1}$$

where B is an N by K matrix of zeros and ones with elements b_{ij} indicating the choice of product j by respondent i. Both ϕ_N and ϕ_K are, respectively, N by N and K by K null matrices. Let D be a diagonal matrix of order $N+K$ with diagonal elements

$$d_{ii} = \sum_{j=1}^{N+K} e_{ij}, \tag{A2}$$

where e_{ij} indicates the (i, j) element of the matrix E in Equation (A1). For $i \le N$, d_{ii} represents the number of products chosen by person i, and for $N+1 \le i \le N+K$, d_{ii} is the number of people choosing product $i-N$. Finally, let x_r be a column vector of $N+K$ coordinates for the N respondents (first N rows) and K products (last K rows) of the rth dimension of the joint space.

Using this notation, Levine's (1979) procedure solves the eigenvalue problem

$$\lambda_r x_r = D^{-1} E x_r, \tag{A3}$$

where λ_r and x_r are the rth eigenvalue and eigenvector, respectively, of the nonsymmetric matrix $D^{-1}E$. In words, this states that a customer's coordinate on dimension r of the map must be proportional (through λ_r) to the centroid of the products he/she selected. Simultaneously, a product's coordinate on dimension r of the map must be proportional (through λ_r) to the centroid of the customers who selected the product.

Appendix B. Multivariate Logistic Distribution

The multivariate logistic distribution describes the probabilities associated with a vector of N binary responses $Z = [Z_1, Z_2, \ldots, Z_N]$. Each of these binary responses can take on two values (zero or one) without restriction. That is, the fact that Z_c has taken on particular value does not restrict the state space of any other variable Z_c. For this reason, there exist 2^N possible values of the Z vector: any string of length N containing only zeros and ones is allowed.

Let $X = [X_1, X_2, \ldots, X_N]$ be a realization of Z. Then, the multivariate logistic distribution of Cox (1972) is defined as

$$\Pr(Z = X) = \exp\Big\{G(X)\Big\} \Big/ \sum_{X^*} \exp\{G(X^*)\Big\}, \tag{B1}$$

where $G(X) = \alpha \sum_c X_c + \sum_{c<c^*} \theta_{cc^*} X_c X_{c^*}$, and the summation in the denominator runs over all 2^N possible values of the Z vector. Here, the notation c^* denotes a customer different than c, and X^* denotes a possible realization of Z (not necessarily X). The cross-effect coefficients θ_{ij} are symmetrical in i and j due to the fact that the correlations among the Zs are determined by these parameters.

Given the structure of Equation (B1), it is easily seen that the full conditional distributions of the multivariate logistic distribution have the form

$$\Pr(Z_c = 1 \mid Z_{c^*} \text{ for } c \ne c^*) = \left[1 + \exp\left[-\left\{\alpha + \sum_{c^* \ne c} \theta_{cc^*} Z_{c^*}\right\}\right]\right]^{-1}, \tag{B2}$$

where again we note that the cross-effect θ_{cc^*} parameters are symmetrical. Equation (B2) is called an AL regression model because we use a logit function and the realizations of other Zs to predict a particular Z_c. Although (B1) implies (B2), it is not clear from simple inspection that the converse is also true. However, using results found in Besag (1974), it can be shown that (B2), along with the symmetry of the cross-effect parameters θ_{cc^*}, implies that the Zs have the multivariate logistic distribution in (B1). In our research, we constrain the θ_{cc^*} parameters to be proportional to the inverse of an exponential function of the distances between customers on the pick-any map (see Equation (3)).

References

Aerts, M., G. Claeskens. 1999. Boostrapping pseudolikelihood models for clustered binary data. *Ann. Inst. Statist. Math.* **51** 515–560.

Andrews, R. L., A. Ansari, I. S. Currim. 2002. Hierarchical Bayes versus finite mixture conjoint analysis models: A comparison of fit, prediction, and partworth recovery. *J. Marketing Res.* **39**(1) 87–98.

Ansari, A., C. F. Mela. 2003. E-Customization. *J. Marketing Res.* **40**(2) 131–145.

Ansari, A., S. Essegaier, R. Kohli. 2000. Internet recommendation systems. *J. Marketing Res.* **37**(3) 363–375.

Ariely, D., J. G. Lynch Jr., M. Aparicio IV. 2004. Learning by collaborative and individual-based recommendation agents. *J. Consumer Psych.* **14**(1–2) 81–95.

Arnold, B. C., D. Strauss. 1991. Pseudolikelihood estimation: Some examples. *Sankhya*: *Indian J. Statist.* **53**(Series B, Part 2) 233–243.

Arnold, B. C., E. Castillo, J. M. Sarabia. 2001. Conditionally specified distributions: An introduction. *Statist. Sci.* **16**(3) 249–274.

Balabanovic, M., Y. Shoham. 1997. Fab: Content-based, collaborative recommendation. *Comm. Assoc. Comput. Machinery* **40**(3) 66–72.

Banerjee, S., B. P. Carlin, A. E. Gelfand. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Monographs on Statistics and Applied Probability 101. Chapman & Hall/CRC, Boca Raton, FL.

Besag, J. 1974. Spatial interaction and the statistical analysis of a lattice system. *J. Roy. Statist. Soc., Ser. B* (*Methodological*) **36** 192–236.

Bodapati, A. 2008. Recommendation systems with purchase data. *J. Marketing Res.* **45**(1).

Bronnenberg, B. J., V. Mahajan. 2001. Unobserved retailer behavior in multimarket data: Joint spatial dependence in market shares and promotion variables. *Marketing Sci.* **20**(3) 284–299.

Bronnenberg, B. J., C. Sismeiro. 2002. Using multimarket data to predict brand performance in markets for which no or poor data exist. *J. Marketing Res.* **39**(1) 1–17.

Cox, D. R. 1972. The analysis of multivariate binary data. *Appl. Statist. (J. Roy. Statist. Soc., Ser. C)* **21**(2) 113–120.

Cressie, N. A. C. 1993. *Statistics For Spatial Data*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York.

Gershoff, A. D., P. M. West. 1998. Using a community of knowledge to build intelligent agents. *Marketing Lett.* **9**(1) 79–91.

Gruca, T. S., D. Sudharshan, K. R. Kumar. 2002. Sibling brands, multiple objectives and responses to entry: The case of the Marion retail coffee market. *J. Acad. Marketing Sci.* **30**(1) 59–69.

Haining, R. 1990. *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press, Cambridge, UK.

Hastie, T., R. Tibshirani, J. Friedman. 2001. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer-Verlag, New York.

Herlocker, J., J. Konstan, J. Riedl. 1999. An algorithmic framework for performing collaborative filtering. *SIGIR'99: Proc. 22nd Annual Internat. ACM SIGIR Conf. Res. Development Inform. Retrieval*, ACM Press, New York, 230–237.

Holbrook, M. B., W. L. Moore, R. S. Winer. 1982. Constructing joint spaces from pick-any data: A new tool for customer analysis. *J. Customer Res.* **9**(June) 99–105.

Iacobucci, D., P. Arabie, A. Bodapati. 2000. Recommendation agents on the internet. *J. Interactive Marketing* **14**(3) 2–11.

Kamakura, W. A., S. N. Ramaswami, R. K. Srivastava. 1991. Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. *Internat. J. Res. Marketing* **8** 329–349.

Kamakura, W. A., M. Wedel, F. de Rosa, J. A. Mazzon. 2003. Cross-selling through database marketing. *Internat. J. Res. Marketing* **20** 45–65.

Kapteyn, A., S. Van De Geer, H. Van De Stadt, T. Wansbeek. 1997. Interdependent preferences: An econometric analysis. *J. Appl. Econometrics* **12** 665–686.

Kim, Y. S., W. N. Street, G. J. Russell, F. Menczer. 2005. Customer targeting: A neural network approach guided by genetic algorithms. *Management Sci.* **51**(2) 264–276.

Kwak, K. 2004. Investigating the applicability of joint-space mapping to new product consideration and choice prediction. Working paper, Tippie College of Management, University of Iowa, Iowa City, IA.

Levine, J. H. 1979. Joint-space analysis of "pick-any" data: Analysis of choices from an unconstrained set of alternatives. *Psychometrika* **44**(1) 85–92.

Mazis, M. B., O. T. Ahtola, R. E. Klippel. 1975. A comparison of four multi-attribute models in the prediction of consumer attitudes. *J. Consumer Res.* **2**(1) 38–52.

Melville, P., R. J. Mooney, R. Nagarajan. 2002. Content-boosted collaborative filtering for improved recommendations. *Proc. Eighteenth National Conf. Artificial Intelligence*, American Association for Artificial Intelligence, Menlo Park, CA, 187–192.

Molenberghs, G., G. Verbeke. 2005. *Models for Discrete Longitudinal Data*. Springer, New York.

Moon, S. 2003. Spatial choice models for product recommendations. Doctoral thesis, Department of Marketing, Tippie School of Business, University of Iowa, Iowa City, IA.

Mooney, R. J., L. Roy. 2000. Content-based book recommending using learning for text categorization. *Proc. Fifth ACM Conf. Digital Libraries*, ACM Press, New York, 195–204.

Moore, W. L., R. S. Winer. 1987. A panel-data based method for merging joint space and market response function estimation. *Marketing Sci.* **6**(1) 25–42.

Pazzani, M. J. 1999. A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Rev.* **13** 393–408.

Rossi, P. E., R. E. McCulloch, G. M. Allenby. 1996. The value of purchase history data in target marketing. *Marketing Sci.* **15**(4) 321–340.

Russell, G. J., A. Petersen. 2000. Analysis of cross category dependence in market basket selection. *J. Retailing* **76**(3) 367–392.

Schafer, J. B., J. A. Konstan, J. Riedl. 2001. E-commerce recommendation applications. *Data Mining Knowledge Discovery* **5** 115–153.

Solomon, M. R., B. Buchanan. 1991. A role-theoretic approach to product symbolism: Mapping a consumption constellation. *J. Bus. Res.* **22** 95–109.

Soucek, B. 1991. *Neural and Intelligent Systems Integration*. John Wiley & Sons, New York.

Strauss, D., M. Ikeda. 1990. Pseudolikelihood estimation for social networks. *J. Amer. Statist. Assoc.* **85**(March) 204–212.

Yang, S., G. M. Allenby. 2003. Modeling interdependent consumer preferences. *J. Marketing Res.* **40**(3) 282–294.