Dynamic Scoring of Customers using Learning Spatial Choice Models

Wolfgang Jank and P. K. Kannan *

June 20, 2005

Abstract

Scoring models are extensively used in CRM applications. However, while most scoring models are static in nature, the quickly changing environment in which firms operate often demands dynamic models that are able to adapt to that change. Moreover, scoring models are often used in environments that span vast and diverse geographical regions. The online channel is only one example where the underlying environment can be geographical extremely diverse. This calls for models that can take the geographical diversity into account: spatial models. In this work we propose spatial models of choice that adapt to dynamically changing environments. These *learning spatial choice models* incorporate new information as it becomes available and are therefore superior over static models. We estimate the model using a version of the online EM algorithm. We apply the learning spatial model of choice to an online publishing firm's data on customer choice of print versus PDF and show how the scoring model can be useful in setting targeted e-coupon promotions or dynamic pricing application.

Key Words: Dynamic Scoring, Spatial Model, Mixed Logit, Monte Carlo, EM Algorithm, Krieging, Dynamic Pricing, Digital Products, Internet

*Wolfgang Jank is Assistant Professor of Decision and Information Technology, and P. K. Kannan is Associate Professor of Marketing, The Robert H. Smith School of Business, University of Maryland, College Park, MD 20742. The authors name are listed in alphabetical order and both have contributed equally to the paper. Please address all correspondence to Wolfgang Jank at the above address; e-mail: wjank@rhsmith.umd.edu; phone 301-405-1118; fax 301-405-8655

1 Introduction

Statistical scoring models are extensively used in CRM applications for ranking customers/ potential customers/ service calls in selection (and retention) of appropriate customers/ service calls for firms. Most of these scoring models are used in a static manner where the models are calibrated on a set of historical data, validated and then applied for scoring purposes. In environments that are relatively static over the application duration, such as credit scoring, the models work reasonably well. However, in dynamic environments, where economy, competition, and consumer preferences change fast, these models have to be periodically updated depending on the need in order to be useful for accurately predicting the dependent variable of interest. For example, in online environments where competitive offerings can change very quickly, firms and models have to be nimble in reacting to, say, price changes for products and services sold online. Similarly, if the scoring model is being used for prioritizing calls arriving at a call center for services such mobile services, ISPs, or ASPs made available over a wide geographical region, it has to take into account changes that arise in local geographical environments (service outage in certain areas, for example) within the last hour or two to be effective. This can be very critical for services that boast high QoS (quality of service) scores. In such cases, scoring models have to incorporate such changes almost dynamically in order to be useful.

In this paper we propose a dynamic scoring model to take into account the fast changing environment characterizing the scoring applications, using learning spatial choice models. Our development of the dynamic scoring model is based on two real life applications which require a dynamic model that "learns" as more data becomes available. Both applications are set in the online environment. The first application is an online mortgage application. The mortgage firm uses the online channel to solicit mortgage refinance applications from potential customers using banner ads and ads on search engines. Customers fill out basic information such as address of the home, income levels, age, and age of the home, and contact information, at the website. These applications are later reviewed and customers are contacted over phone to explore the lead further. The firm, however, needs a scoring model to choose the most promising application to make the sales most efficient. The firm has identified geographical location as an important factor in predicting sales success with a given customer – those geographical areas where property values have risen significantly and/or houses at the appropriate age provide good prospects for refinancing of homes. In this instance, spatial data can act as a proxy for many geography related variables - income, education, age, house size, property values, etc. – and capture some of the variances in customer sales success due to these variables. However, given that local refinancing firms are also quite active in certain local geographical areas and are using different, continuously changing tactics to attract new customers, any scoring model that is developed cannot be just static but has to take into account changes in the environment as they evolve, updating the model with new data as it becomes available. This is strong motivation for the development of a scoring model that not only considers spatial elements but also the dynamic nature of the market.

The second application involves an online publishing firm that sells titles in both print form and electronic (PDF) form. The firm has priced the relative forms of content based on pricing experiments in the past, which considered the variations of preferences for print versus PDF across different geographical regions (due to how tech savvy the customers are, how technology ready they are, whether they are high income or highly educated or not, and whether or not broadband access has penetrated the local market to a significant extent). However, these spatial factors change over time fairly quickly. In addition, social contagion and word of mouth effects in local markets are quite unpredictable and may change quickly in a matter of only days. If this is the case, just considering the spatial effects in a static model to compute optimal prices for content may not be sufficient. Using a dynamic learning model, any dynamics in spatial markets can be immediately incorporated in the scoring model and effective prices for the books in print and PDF form can be changed by providing online coupons for customers targeted in specific geographical areas. Thus such scoring models can be very useful for dynamic pricing applications for online services and firms.

The obvious question in using spatial model in online contexts is "why is it useful". In general, spatial data supplements the data available about online consumers. Obviously, online context helps to collect a lot of details on individual consumers. However, this occurs over time as the volume of transactions/interactions grow. During such times when available data is limited, spatial data can be used. All we need is zip-code info which can be gathered from transaction info or inferred (fairly accurately) from IP addresses. In some cases, there might be privacy concerns which may limit the amount of information that can be gathered from customers directly. In such cases, online retailers can use the data available through transactions (zip-code, for example) as proxy data to account for customer differences on many dimensions. Even when substantial information about customers are available, one can still use spatial data to account for unobserved variables, variables that managers have failed to account for, which may provide additional insights into customer choice. Spatial data can be especially important for multi-channel firms which use traditional channels along with the Internet channel to interact with customers. For example, many online newspapers collect data on readers' zip code (even though the content is free) to determine the cannibalization impact on print. For our own application, understanding customers behavior in different geographical regions can help in promoting online channel and specific forms of products using targeted e-coupons.

Even if zip-code level data is available, spatial models can still be useful. Most of the variables the zip-code data represent come from the census and other sources and include income, age, and other demographic information. In the online context, while rich data is available on customer behavior (click-throughs, previous purchases, prices paid, coupons redeemed etc), demographic information may not be easily available. In many cases, online firms buy such data from outside sources. It is data such as these that our zip-code data may act as proxy for. So it might not be so superfluous and may have value, especially, by providing additional predictive power regarding customer choices. Most online firms buy data from outside sources and append to their database - there is a lot of missing info on the details such as income, age, and other demographics - so the spatial model can be a good supplement in such cases. Finally, parsimony in modeling spatial data as we do, is another great advantage.

There has been a strong recent interest in incorporating spatial information into traditional statistical models. Successful applications include Bronneberg & Sismeiro (2002), Hofstede, Wedel & Steenkamp (2002), Mittal, Kamakura & Govind (2004) and Rust & Donthu (1995). However, while most consider situations where the data is continuous, Jank &

Kannan (2005) propose a model for spatial choice. In the following we build upon their work and derive a "Learning Spatial Choice Model."

In the next few sections, we describe our model of spatial choice and show how learning can be incorporated as new data arrives. Then, we apply the learning spatial model of choice to an online publishing firm's data on customer choice of print versus PDF and show how the scoring model can be useful in setting targeted e-coupon promotions or dynamic pricing application.

2 Spatial Model of Customer Choice

2.1 Description of Choice Model

Our spatial model allows for geographical dependence of consumer choices. We model choices via a multinomial logit model. We allow for the possible spatial dependence of choice decisions by including a stochastic parameter vector with suitably chosen geographically varying correlation structure. This is similar to the spatial model proposed by Jank & Kannan (2005).

Let $\mathbf{z}_i = (z_{i1}, z_{i2}), i = 1, \ldots, N$, denote the spatial coordinate or location of the observed response y_i. Our model assumes that the response variable takes on only one of J values, $y_i \in \{1, 2, \ldots, J\}$, and that, conditional on $\mathbf{z}_i$, the y_i's are independent realizations of a multinomial random variable; that is

$$y_i|\mathbf{z}_i \sim \text{Multinomial}(\pi_{i1}, \pi_{i2}, \ldots, \pi_{iJ}), \tag{1}$$

where $\pi_{ij} = \text{Prob}(y_i = j|\mathbf{z}_i)$ is the probability of choosing category $j, (j = 1, \ldots, J)$, conditional on location $\mathbf{z}_i$. Let J denote the *baseline* category. We model the logit of π_{ij} as

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \mathbf{x}_i^T \boldsymbol{\beta}_{ij}, \quad j = 1, \ldots, J-1, \tag{2}$$

where $\mathbf{x}_i$ is $p \times 1$ a vector of known covariates and $\boldsymbol{\beta}_{ij}$ is a $p \times 1$ vector of unknown logit parameters associated with category j. We assume that the $\boldsymbol{\beta}_{ij}$'s vary across geographical regions according to a multivariate normal distribution that is centered at $\tilde{\boldsymbol{\beta}}_j$ and has covariance matrix $\boldsymbol{\Sigma}$. One of the modeling goals is to allow for a higher correlation between parameters $\boldsymbol{\beta}_{ij}$ and $\boldsymbol{\beta}_{i'j}$ if $\mathbf{z}_i$ and $\mathbf{z}_i'$ are in closer geographic proximity.

We achieve this goal by modeling the covariance matrix $\boldsymbol{\Sigma}$ accordingly. In particular, we want the correlation between $\mathbf{z}_i$ and $\mathbf{z}_i'$ to decay with their geographical distance. To that end, we make standard assumptions (e.g. Cressie, 1993) like homogeneity of the variance, and second-order stationarity and isotropy which ensures translation and rotation invariance of the correlation structure. As a result we can write the covariance between any two points $\mathbf{z}_i$ and $\mathbf{z}_i'$ as

$$\text{Cov}(\boldsymbol{\beta}_{ij}, \boldsymbol{\beta}_{i'j}) = \tau_j^2 + \sigma^2 r(d_{i,i'}; \boldsymbol{\alpha}), \tag{3}$$

where τ_j^2 captures pure modeling error and is often referred to as the *nugget effect*. Notice that we allow τ_j^2 to be alternative-specific. Spatial association is captured by the correlation function $r(\cdot)$, which is a function of the (Euclidian) distance $d_{i,i'} = \|\mathbf{z}_i - \mathbf{z}_i'\|$ between location i and i', scaled by σ^2 and dependent on a vector of spatial parameters $\boldsymbol{\alpha}$.

2.2 Modeling Spatial Correlation

A variety of different correlation functions can be used. We adopt a very flexible class of correlation functions called the *Matérn* family, which defines

$$r(d;\boldsymbol{\alpha}) = \frac{(d/\alpha_1)^{\alpha_2}}{2^{\alpha_2-1}\Gamma(\alpha_2)} K_{\alpha_2}(d/\alpha_1), \tag{4}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T$, $\alpha_1, \alpha_2 > 0$, and K_α is the modified *Bessel* function of (possibly fractional) order α. The parameter α_1 is often referred to as the spatial scale parameter while α_2 is the shape parameter. Setting $\alpha_2 = 0.5$ recovers, as a special case, the *exponential* correlation function

$$r(d;\alpha_1) = \exp(-\alpha_1 d), \tag{5}$$

which implies that the correlation between two observations decays exponentially with the geographical distance between the associated two observations. For more modeling options see e.g. Cressie (1993).

3 Learning via Expectation-Maximization

3.1 Monte Carlo EM for Spatial Choice

Let $\mathbf{y} = (y_1, \ldots, y_N)$ denote the vector of the observed choice responses. We can frame the spatial choice model in the context of the well-known Expectation-Maximization (EM) algorithm by assuming that $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_N)$ is a vector of (unobserved) random effects. In particular, let $\mathbf{u}_i = (\mathbf{u}_{i1}, \ldots, \mathbf{u}_{i(J-1)})$ where $\mathbf{u}_{ij}$ follows a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}$. Then we can re-write equation (2) as $\log(\pi_{ij}/\pi_{iJ}) = \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}_j + \mathbf{x}_i^T \mathbf{u}_{ij}$. Let $\boldsymbol{\theta}$ be the vector of all parameters. We learn $\boldsymbol{\theta}$ via the EM algorithm in the following way.

The EM algorithm (Dempster, Laird & Rubin, 1977) is an iterative learning method. Let $\boldsymbol{\theta}^{(l-1)}$ denote its current parameter value. Then, in the l^{th} iteration of the algorithm, the E-step of the method computes the expectation $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l-1)}) = E\left[\log f(\mathbf{y}, \mathbf{u}; \boldsymbol{\theta})|\, \mathbf{y}; \boldsymbol{\theta}^{(l-1)}\right]$. In the context of spatial choice, this expectation involves an analytically intractable integral but we approximate it readily via Monte Carlo. Let

$$\tilde{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l-1)}) = \frac{1}{m_t} \sum_{k=1}^{m_t} \log f(\mathbf{y}, \mathbf{u}^{(k)}; \boldsymbol{\theta}), \tag{6}$$

where $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(m_t)}$ are simulated from the conditional distribution $f(\mathbf{u}|\mathbf{y}; \boldsymbol{\theta}^{(l-1)})$. This gives rise to the Monte Carlo EM (MCEM) algorithm. In the M-step, we maximize (6) to obtain the l^{th} update, $\boldsymbol{\theta}^{(l)}$. We repeat this process until the parameter updates converge.

There are several important issues when implementing MCEM. The Monte Carlo sample size m_t should be increased successively as the algorithm moves along (see Caffo, Jank & Jones, 2005, for an automated selection of m_t). Also, since MCEM is stochastic in nature, the stopping criterion has to be chosen with care (see Booth, Hobert & Jank for more discussion). In every iteration, we have to simulate $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(m_t)}$ from $f(\mathbf{u}|\mathbf{y}; \boldsymbol{\theta}^{(l-1)})$. Jank

(2004) proposes an efficient simulation approach based on importance sampling and Quasi-Monte Carlo variance-reduction techniques. Once the parameter estimates are obtained, we can calculate their standard errors from the output of MCEM's final iteration using Louis' method (see Louis, 1982).

3.2 Spatial Prediction with MCEM

One of the major appeals of spatial models is their ability to predict for locations where no data is observed. This is typically done by borrowing information from surrounding areas for which information is available and this is often referred to as *krieging*. Spatial prediction can be incorporated naturally within the framework of the MCEM algorithm. Spatial prediction consist of two steps: first, one *estimates* the random effects from observed locations via simulation. Then, using these estimated random effects, one *predicts* the random effects for unobserved locations and uses those for predicting a customer's choice.

Given the parameter vector $\hat{\boldsymbol{\theta}}$ from the MCEM output, one estimates the random effects by approximating the minimum mean squared error estimate of $\mathbf{u}$. This value is given by the conditional expectation $E[\mathbf{u}|\mathbf{y};\hat{\boldsymbol{\theta}}]$. While this expectation cannot be computed in closed form, we can approximate it via $\hat{\mathbf{u}} = (\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_N) = \frac{1}{m}\sum_{k=1}^{m} \mathbf{u}^{(k)}$, where $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(m)}$ are simulated from $f(\mathbf{u}|\mathbf{y};\hat{\boldsymbol{\theta}})$. In other words, $\hat{\mathbf{u}}_i$ is the estimated random effect corresponding to location $\mathbf{z}_i$.

Using the $\hat{\mathbf{u}}_i$'s, we can predict a customer's choice. Let $\mathbf{z}_*$ denote a location for which no data has been observed. Let $\mathbf{u}_*$ denote the corresponding random effect. The linear predictor $\hat{\mathbf{u}}_*$ of $\mathbf{u}_*$ is given by $\hat{\mathbf{u}}_* = E[\mathbf{u}_*|\mathbf{y}] = \mathbf{C}_*\hat{\mathbf{u}}$, where the coefficient matrix $\mathbf{C}_*$ is such that $E[\mathbf{u}_*|\mathbf{u}] = \mathbf{C}_*\mathbf{u}$ (for more details see Zhang, 2002). Given the predicted random effect $\hat{\mathbf{u}}_*$, we can calculate the predicted choice probability for an unobserved location $\mathbf{z}_*$ via

$$\pi_{*j} = \exp(\mathbf{x}_*^T\hat{\boldsymbol{\beta}}_j + \mathbf{x}_*^T\mathbf{u}_{*j})/[1 + \sum_{j=1}^{J-1}\exp(\mathbf{x}_*^T\hat{\boldsymbol{\beta}}_j + \mathbf{x}_*^T\mathbf{u}_{*j})]. \tag{7}$$

3.3 Online learning of the choice model

In many situations information becomes available only successively. For instance, new customers arrive at an online business over the course of weeks and months. In order to make up-to-date predictions, every new piece of information has to be incorporated into the choice model as it becomes available. In the following we propose an online version of the EM algorithm to update the spatial model in real-time. We consider two different scenarios: a short-term and a long-term scenario. In the short-term we assume that the model parameters remain roughly constant. However, this assumption may not be valid in the long-term and we adjust the method accordingly.

3.3.1 Online Learning in the Short-Term

Let $\mathbf{y}_{t-1} = (y_1, \ldots, y_{t-1})$ denote the training data observed over the previous t-1 time periods. We initialize the model by applying MCEM to $\mathbf{y}_{t-1}$ and obtain an initial estimate $\boldsymbol{\theta}_{t-1}$. The goal is to update $\boldsymbol{\theta}_{t-1}$ as new information arrives.

Let $t, t+1, t+2, t+3, \ldots,$ denote the times at which new information becomes available and let y_t denote the information at time t. Since we assume that in the short-term the model parameters remain constant, statistical reasoning suggests that we obtain a more accurate model by incorporating all of the available information. Let $\mathbf{y}_t = (\mathbf{y}_{t-1}, y_t)$ denote the entire data available at time t and let $\boldsymbol{\theta}_t$ denote the corresponding parameter estimate via MCEM.

We use this information to predict the next observation in the following way. A new customer arrives at time $t+1$. We predict this customer's choice using spatial prediction based on $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_t$ and the resulting choice probability in equation (7). Once we observe the customer's true choice y_{t+1}, we incorporate it into the training data and learn the model parameters all over again.

3.3.2 Online Learning in the Long-Term

In the long-term, customer preferences may change. The choice model has to be adaptive to reflect that change. In the following we describe an online learning approach that can account for that change.

If customer preferences change over time, then earlier information is less relevant and should be discounted. To that end, let $\mathbf{y}_{t,L} = (y_{t-L}, y_{t-L+1}, \ldots, y_t)$ denote the information from only the last L customers. Let $\boldsymbol{\theta}^*$ be MCEM's estimate based on (the reduced) training data $\mathbf{y}_{t,L}$. We update the parameter gradually by discounting the influence of the previous parameter value

$$\boldsymbol{\theta}_t = \gamma\boldsymbol{\theta}^* + (1-\gamma)\boldsymbol{\theta}_{t-1}, \tag{8}$$

where the discounting parameter is $0 < \gamma < 1$. This online version of EM shares similarities with the adaptive EM algorithm of Ng & McLachlan (2004). One can calibrate the values of L and γ such that prediction accuracy is maximal on the training data.

4 Experiment

In order to investigate the performance of the online learning algorithm, we simulated data from our spatial choice model. We considered two scenarios: one in which the parameters remained constant across all simulations and another in which the parameter values gradually changed over time.

We consider the situation of an online book publisher who sells its titles in both print and PDF format. A customer's preference for either format depends, among other things, on the product's price and other, geographically varying factors like broadband access or tech-savviness. This gives rise to a model of the form (2), (3) and (4) where $J = 2$ (print vs. PDF), $\tilde{\boldsymbol{\beta}}_j$ captures the (geographically varying) price sensitivities, and τ_j^2, σ^2 and $\boldsymbol{\alpha}$ describe the spatial correlation.

4.1 Short-term Learning - Constant Parameters

In the first scenario we assume that the model parameters $\tilde{\boldsymbol{\beta}}_j$, τ_j^2, σ^2 and $\boldsymbol{\alpha}$ remain constant over time. We simulate two sets of data from the resulting model: a training data set $\mathbf{y}_{t-1} = (y_1, \ldots, y_{t-1})$ and a validation data set $\mathbf{y}_* = (y_t, y_{t+1}, \ldots, y_T)$. The assumption is that $\mathbf{y}_{t-1}$ contains customer transactions observed in the past while the $\mathbf{y}_*$'s are future observations. The book publisher would like to predict an observation in $\mathbf{y}_*$ based on previous information.

This prediction task can be tackled in two different ways. One way is to calibrate the spatial choice model only on $\mathbf{y}_{t-1}$ and use the resulting static model to predict all observations in $\mathbf{y}_*$. We refer to this as the "batch approach" since the estimation process is done in one batch. The alternative approach is to update the model as new information becomes available. This is the online learning model from Section 3.3.

Table 1 shows the misclassification rates for the online and for the batch approach. While the batch approach misclassifies 77.55% of the validation data, the number is only 44.90% for the online algorithm. It is interesting to note that over 81% of all customers in the validation set chose the print format. On the other hand, the training set had only 47% print preferences, a significantly smaller proportion. The batch approach has a hard time adjusting for that. The batch method is calibrated entirely on the training data and cannot "learn" from mistakes. This is different for the online learning algorithm and consequently it results in a better forecasting accuracy.

	Online Predicted		Batch Predicted		
True	PDF	Print	PDF	Print	Total
PDF	6.12%	12.24%	16.33%	2.04%	18.37%
Print	32.65%	48.98%	75.51%	6.12%	81.63%
Total	38.78%	61.22%	91.84%	8.16%	100.00%

Figure 1: *Constant Parameters:* Misclassification matrix for the Online Learning Algorithm and for the Batch Algorithm. Total misclassification rates are 44.90% and 77.55%, respectively.

4.2 Long-term Learning - Time-Varying Parameters

Learning is especially important when the population changes. For instance, customers' preferences for PDF may change over time as broadband becomes available in more and more areas, or as a the "fear of new technology" becomes smaller due to media and news coverage. In the spatial choice model, changing customers preference is manifested in time-varying parameters. We again simulate a training and validation set from our model. This time, however, we let the parameters $\tilde{\boldsymbol{\beta}}_j$, τ_j^2, σ^2 and $\boldsymbol{\alpha}$ change gradually over time. We again benchmark the online learning algorithm against the batch approach. Table 2 shows the results.

We can see that the misclassification rate for the batch algorithm is huge compared to the online learner. In fact, the batch method classifies every single observation as "print." The reason for that is the time-changing parameter. We changed the parameters in a way that while only about 20% of the customers in the training set preferred PDF, this number is about 70% in the validation set. Since the batch method is calibrated on the training set only, its predictions are biased towards the print format. The online learning algorithm on the other hand adapts to the changing environment.

	Online Predicted		Batch Predicted		
True	PDF	Print	PDF	Print	Total
PDF	48.98%	20.41%	0%	69.39%	69.39%
Print	4.08%	26.53%	0%	30.61%	30.61%
Total	46.94%	53.06%	0%	100.00%	100.00%

Figure 2: *Time-Varying Parameters:* Misclassification matrix for the Online Learning Algorithm and for the Batch Algorithm. Total misclassification rates are 24.49% and 69.39%, respectively.

5 Empirical Application

The simulation experiment has provided us promising results regarding the effectiveness of the learning spatial model in terms of its scoring (prediction) performance. We are currently working the online publishing firm to implement the learning model for dynamic targeting of customer as they enter the website. The steps involved in the process are as follows:

1. Given the current (prevailing) prices of print and PDF forms of the titles in different subject content, the estimated spatial model parameters (applied in a static framework) will be used first to generate purchase probabilities (conversion rates) of customers arriving from different spatial points (inferred using their IP addresses) and interacting with a specific title. (Note that the conversion rates are modelled conditional on customers interacting with a particular title online, and hence they are much higher than 3- 5% unconditional conversion rates that characterize online browsing.)

2. Given new data generated at the website in a period of 4 hours, the learning model will be used to generate the updated purchase probabilities (conversion rates). If the predicted conversion rates are lower than before, then predictions of purchase probabilities for a "percent off" e-coupon promotion will be determined for various levels of e-coupon value (ranging from 5% to 20%). Based on the predicted responses, customers arriving from more price sensitive geographical areas will be targeted with the e-coupon promotion. (A similar scheme of e-coupon price discrimination at website has also been suggested by Dogan, Cheng and Kohler (2004) and Dogan and Cheng (2004). However, our model is superior to theirs in the sense that we have a scoring methodology and it takes into account learning.)

3. The predicted hit rates will be compared with actual hit rates to measure the effectiveness of the learning model and make refinements.

We expect the empirical analysis to be completed by August 2005.

References

Booth, James G, James P Hobert, and Wolfgang S Jank (2001), "A survey of Monte Carlo algorithms for maximizing the likelihood of a two-stage hierarchical model," *Statistical Modelling*, 1, p.333-349.

Bronnenberg, Bart J. and Catarina Sismeiro (2002), "Using Multimarket Data to Predict Brand Performance in Markets for Which No or Poor Data Exist," *Journal of Marketing Research*, Vol. 39, February, p. 1-17.

Caffo, Brian S, Wolfgang S Jank, and Galin L Jones (2003), "Ascent-Based Monte Carlo EM", *Journal of the Royal Statistical Society, Series B.*

Cressie, Noel AC (1993), *Statistics for Spatial Data, Revised Edition*, Wiley, New York.

Dempster, Art P, Nan M Laird, and Donald B Rubin (1977), "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society B*, 39, p. 1-22

Dogan, Kutsal and Hsing Kenneth Cheng (2004) "One-to-one Marketing with Internet Coupons: Spatial Models of Price Discrimination and Targeting", submitted to *Journal of Management Information Systems*

Dogan, Kutsal, Hsing Kenneth Cheng and Gary J. Koehler (2004) "A Spatial Model of Price Discrimination with Coupons", in preparation.

Hofstede, Frenkel Ter, Michel Wedel, and Jan-Benedict E. M. Steenkamp (2002), "Identifying Spatial Segments in International Markets," *Marketing Science*, 21 (2), Spring, p. 160-177.

Jank, Wolfgang S and PK Kannan (2005), "Understanding geographical markets of online firms using spatial models of customer choice", *Marketing Science*, (forthcoming).

Jank, Wolfgang S (2004), "Quasi-Monte Carlo Sampling to Improve the Efficiency of Monte Carlo EM", *Computational Statistics and Data Analysis*,48, p.685701.

Louis, Thomas A (1982), "Finding the observed information matrix when using the EM algorithm, " *Journal of the Royal Statistical Society B*, 44, p.226-233.

Mittal, Vikas, Wagner Kamakura and R. Govind (2004), "Geographic Patterns In Customer Service and Satisfaction: An Empirical Investigation," *Journal of Marketing*, (forthcoming).

Ng, S.K. and Geoff J. McLachlan (2004). "An intelligent system for on-line prediction of hospital resource utilization: an incremental EM-based learning approach", in preparation.

Rust, Roland T. and Naveen Donthu (1995), "Capturing Geographically Localized Misspecification Error in Retail Store Choice Models," *Journal of Marketing Research*, Vol. 32 (February), 103-110.

Zhang, Hao (2002), "On estimation and prediction for spatial generalized linear mixed models", *Biometrics*, 58, p.129-136