.... RESEARCH NOTES AND COMMUNICATIONS

ASIM ANSARI, SKANDER ESSEGAIER, and RAJEEV KOHLI*

Several online firms, including Yahoo!, Amazon.com, and Movie Critic, recommend documents and products to consumers. Typically, the recommendations are based on content and/or collaborative filtering methods. The authors examine the merits of these methods, suggest that preference models used in marketing offer good alternatives, and describe a Bayesian preference model that allows statistical integration of five types of information useful for making recommendations: a person's expressed preferences, preferences of other consumers, expert evaluations, item characteristics, and individual characteristics. The proposed method accounts for not only preference heterogeneity across users but also unobserved product heterogeneity by introducing the interaction of unobserved product attributes with customer characteristics. The authors describe estimation by means of Markov chain Monte Carlo methods and use the model with a large data set to recommend movies either when collaborative filtering methods are viable alternatives or when no recommendations can be made by these methods.

Internet Recommendation Systems

Recommendation systems provide a type of mass customization that is becoming increasingly popular on the Internet. Search engines such as Yahoo! and Alta Vista use them to recommend relevant documents on the basis of user-supplied keywords. The *Los Angeles Times* allows online news customization. Amazon.com and barnesandnoble.com recommend books and movies on the basis of the preferences of their other customers. Such customization ostensibly decreases the search effort for users. It also promises a firm greater customer loyalty, higher sales, more advertising revenues, and the benefit of targeted promotions.

Current customization systems fall into two classes that use different information sources to make recommendations. The first class comprises collaborative filtering, which mimics word-of-mouth recommendations. Operationally, these methods predict a person's preferences as a linear, weighted combination of other people's preferences. Notable commercial implementations of collaborative filtering are offered by Net Perceptions, Likeminds, and the now defunct Firefly. The second class, known as content filtering, makes recommendations on the basis of consumer preferences for product attributes. The available commercial systems offered by PersonalLogic, Frictionless Commerce, and Active Research use self-explicated importance ratings and/or attribute trade-offs to make their recommendations.

Both types of filtering methods have limitations. Collaborative filtering needs dense data sets; can be used only when at least a few people have evaluated a product; does not reflect uncertainty in predictions; and provides few, if any, reasons for a recommendation. Attribute-based systems can recommend entirely new items but do not necessarily incorporate the information in preference similarity across individuals. Similar to collaborative filtering, these methods also cannot make recommendations for people who provide no preference information. And though it is desirable in online situations to minimize the amount of data collected from a person, little is known about the trade-off

Asim Ansari is an associate professor (e-mail: maa4@columbia.edu), Skander Essegaier is a doctoral candidate (e-mail: se57@columbia.edu), and Rajeev Kohli is a professor (e-mail: rk35@columbia.edu), Graduate School of Business, Columbia University. The authors thank Digital Equipment Corporation for generously providing the data.

between predictive accuracy and the amount of data collected from a person for this class of recommendation system. Finally, to our knowledge, there are no published comparisons within or across the two types of recommendation systems, nor is there a comparison with uncustomized recommendations that is obtainable, for example, from an attribute-based regression that pools data across individuals.

As is clear from this description, collaborative and content filtering methods use one or two types of information. However, there are at least five information sources that can be used for making recommendations: (1) a person's expressed preferences or choices among alternative products, (2) preferences for product attributes, (3) other people's preferences or choices, (4) expert judgments, and (5) individual characteristics that may predict preferences. A good recommendation system should be able to use any or all of these five types of information, potentially making better recommendations as more information becomes available. In other words, a method should be able to integrate alternative information sources by means of nested specifications that allow for predictions based on information subsets. It should also provide estimates of its accuracy, explain reasons behind recommendations, and incorporate dynamic learning in the sense that as more information becomes available for certain people, it should make better recommendations for those and possibly other people.[1]

We use a hierarchical Bayesian approach to design a recommendation system. Similar to the models described by Allenby and Ginter (1995) and Rossi and Allenby (1996), ours allows unobserved heterogeneity in consumer preferences. In addition, we introduce the effect of unobserved product heterogeneity on preferences to allow for the introduction of unobserved product attributes, such as holistic customer judgments and product appeal structures. In an online context, accounting for product heterogeneity is crucial because the product/merchant options available to a consumer often change on an ongoing basis. It therefore makes sense to consider alternatives at a certain time as random draws from a suitable distribution. This contrasts with models for supermarket purchases by a panel of households, for which a fixed (often small) number of brands sold by a single retail chain is best represented by a fixed-effects model.

In the following section, we briefly discuss recommendation systems; we then describe the set of proposed models and their estimation using Markov chain Monte Carlo (MCMC) methods. In the final section, we present the results of a test of the models using a large, publicly available data set that other researchers have used to make online movie recommendations to consumers.

AN OVERVIEW OF RECOMMENDATION SYSTEMS

Although the idea dates back to Negroponte (1970) and Kay (1984), practical implementations of "intelligent agents" are relatively recent and have been fueled by the successes of online companies such as Firefly and Amazon.com. Early uses, which were not on the Internet, included short-lived in-store kiosks at Blockbuster Video that recommended films on the basis of a member's past rental history (West et al. 1999). This made possible such interesting recommendations as pornographic films to children and Teletubbies programs to grandparents who lived in the same household. Its effect on family well-being is not known. Then there was Magnet (Levy 1993), which claimed to be the first intelligent agent for the Macintosh. Essentially a file manager, it dispatched files into the trash if a user mistyped a destination folder (Foner 1993).

Recommendation systems are agents of the sort used by Blockbuster. Using behavioral or preference information, they filter alternatives and make suggestions to a user. Internet search engines are an example of such content-based systems, as these retrieve documents by means of keywords. In one commonly used system, the frequency of a target word is used to assess a document's relevance, and the relative frequencies of words are used to assess document similarity (Salton and Buckley 1988).

Similar to conjoint analysis, recommendation systems screen attractive alternatives. But whereas conjoint analysis is typically used to screen many products to find a few attractive market options, recommendation systems are models for individual-level prediction that can be useful even if there are a few alternatives. For example, a person deciding among new releases of music, plays, or movies may have only a few choices. Recommendation systems are also eminently suitable for suggesting books, restaurants, dry cleaners, plumbers, physicians, lawyers, financial institutions, and real estate brokers. And it is not just experience or reputation services of this sort for which people may seek recommendations. As anyone who has used *Consumer Reports* and the so-called online product configurators knows, recommendations can be useful for such products as cars and computers, which are considered search goods but which people do not always have the ability or the means to evaluate. Perhaps most important, recommendation systems need to work well with much less information than is generally collected in marketing research studies, in which either a respondent is compensated for participating in a one-time study or choice data are available from customer panels over a relatively long period of time. In contrast, the folklore for online recommendation systems is that most people are averse to answering too many questions before they get recommendations.

Collaborative filtering algorithms were first introduced by Goldberg and colleagues (1992). They are used by the *Los Angeles Times*, *London Times*, CRAYON, and Tango to customize online newspapers; by Bostondine to recommend restaurants in and around Boston; by Sepia Video Guide to make customized video recommendations; by Movie Critic, Moviefinder, and Morse to recommend movies; and by barnesandnoble.com to recommend books. In the simplest case, collaborative filtering predicts a person's preferences as a weighted sum of other people's preferences, in which the weights are proportional to correlations over a common set of items evaluated by two people. More recently, model-based collaborative filtering has been introduced. We refer the reader to Breese, Heckerman, and Kadie (1998) for a description of these alternative implementations, which include Bayesian networks (Heckerman 1996) and finite mixture models (Chien and George 1999). An early assessment of these methods by Breese, Heckerman, and Kadie (1998) is not encouraging on reported predictive criteria.

[1]Lycos has recently acquired a system called WireWise that provides adaptive filtering. The details of its method are described in U.S. Patent #5,983,214.

As already noted, collaborative filtering algorithms have several limitations. First, when data are sparse, the correlations (weights) are based on few common items and therefore are unreliable. Breese, Heckerman, and Kadie (1998) show that prediction performance suffers dramatically in such a situation. Second, collaborative filtering algorithms can be used only when preference data for an item already exists in the database. In other words, these systems cannot handle queries that pertain to new items. For example, most collaborative filtering algorithms cannot help a user who needs to know whether a new movie is good. In such situations, the database has no information about the movie, and the system is therefore unable to process such requests. Third, these methods use ad hoc prediction algorithms, which are not based on a statistical model. Consequently, they do not account for uncertainty, which may be less important for such low-risk purchases as movies and compact discs but can be very important when the stakes are higher for a consumer or company. Fourth, collaborative filtering systems do not explicitly incorporate attribute information, though they are bootstrapped by creating "virtual users" who represent particular tastes (e.g., a virtual action fan who has high ratings for all action movies). The implications of such indirect accounting of product features is not clear. Finally, because collaborative filtering methods are correlational, they provide little explanation for a recommendation, a feature that can be important for building trust and enhancing customer loyalty.

Attribute-based systems allow recommendations for entirely new items but do not necessarily incorporate the information in preference similarity across individuals.[2] Similar to collaborative filtering, these methods cannot make recommendations for people who provide no preference information. The commercially available attribute-based systems appear to collect a large amount of information from respondents, and little is known about their predictive accuracy. Moreover, systems that use neural networks often have difficulty providing explanations for recommendations. To overcome the shortcomings of existing systems, we develop flexible yet simple statistical models, which are described in the next section.

A HIERARCHICAL BAYESIAN RECOMMENDATION SYSTEM

We develop an ensemble of statistical methods, which we estimated using customer ratings on small, idiosyncratic subsets of products. We then use these models to make customized recommendations over holdout items—in our case, new theater releases and video rentals (or rereleases) of older movies. We adopt a regression-based approach and model customer ratings as a function of product attributes, customer characteristics, and expert evaluations. The models we develop differ in how they account for unobserved sources of heterogeneity in customer preferences and product appeal structures. To the extent that parameter estimates reflect causal preference structures, they enable us to tell not only what people may like but also why they may react in the predicted manner.

Customer Heterogeneity

The database consists of ratings provided by customers for many different movies. Customers differ in the number of movies they rate, which yields an unbalanced data set. Let i = 1 to I represent customers and j = 1 to J represent movies. Customer i provides ratings for n_i movies in the database; let $M_i = \{j_1, j_2, \ldots, j_{n_i}\}$ denote the index set of the n_i movies rated by customer i. Let r_{ij} represent the rating given by customer i for movie j, where $j \in M_i$. The total number of ratings across all customers is given by $N = \Sigma_{i=1}^{I} n_i$.

The observations for each customer are used to specify a customer-level regression model:

$$(1) \qquad r_{ij} = \mathbf{w}_j \boldsymbol{\beta}_j + e_{ij},\ e_{ij} \sim N(0, \sigma^2),$$

where $j \in M_i$, $\mathbf{w}_j$ is a vector of movie attributes (genre and expert ratings) for movie j, and $\boldsymbol{\beta}_i$ is a vector of parameters that represent the preference structure for customer i.

If the database contains many observations for each customer, we can in principle estimate the preceding regression model for each customer. In many situations, however, the database is sparse, and only a few observations are available for some customers. Therefore, we cannot perform separate regressions for each customer. We can, however, use a hierarchical Bayesian approach that adequately pools information across customers to make inferences that pertain to a specific customer. In this approach, a continuous mixture distribution is used to describe how the individual-level parameters in Equation 2 vary across the customers in the population. The population model that accounts for both observed and unobserved sources of heterogeneity is

$$(2) \qquad \boldsymbol{\beta}_i = \mathbf{z}_i \boldsymbol{\mu} + \boldsymbol{\lambda}_i,$$

for i = 1 to I. In Equation 2, $\mathbf{z}_i$ contains the characteristics of customer i, and $\boldsymbol{\lambda}_i$ represents the unobserved customer effect for the ith customer.

The complete model can alternatively be written as

$$(3) \qquad r_{ij} = \mathbf{x}'_{ij}\boldsymbol{\mu} + \mathbf{w}'_j\boldsymbol{\lambda}_i + e_{ij},\ e_{ij} \sim N(0, \sigma^2),\ \boldsymbol{\lambda}_i \sim N(\mathbf{0}, \boldsymbol{\Lambda}),$$

for i = 1 to I and $j \in M_i$. In Equation 3, $\mathbf{x}_{ij}$ is a vector containing all observed movie attributes (i.e., genre variables and expert ratings) and person characteristics and their interactions, and $\mathbf{w}_j$ is a vector that contains the observed movie attributes. The vector $\boldsymbol{\mu}$ represents the fixed effects and documents the influence of observed customer and movie variables and their interactions. The vector $\boldsymbol{\lambda}_i$ contains all the random effects pertaining to the ith customer. The covariance matrix $\boldsymbol{\Lambda}$ provides information about the extent of unobserved heterogeneity in customer preference structures.

Product Heterogeneity

Previous approaches to modeling heterogeneity in marketing (mostly in conjoint and discrete choice contexts) have used data that involve a few products that are well described by observed attributes. In such contexts, differences in customer preference structures primarily contribute to the heterogeneity in the data. In contrast, recommendation systems operate on databases that include ratings on many products. Moreover, as in the case of movies and

[2] Shardanad and Maes (1995) and Balabanović and Shoham (1997) provide examples of systems that combine content and collaborative filtering. Sarwar and colleagues (1998) use filterbots that act like normal users in a collaborative filtering system and rate articles on the basis of certain semantic information.

music, products cannot be described adequately in terms of a few observable attributes. Consumer preferences in such categories are shaped by myriad attributes that interact in intricate ways, which leads to thematic differences that necessitate accounting for these complex yet unobserved product attributes (see Gershoff and West 1998). These unobserved movie attributes lead to differences in product appeal structures. Accommodating these differences among movies becomes crucial in modeling customer ratings. In this section, we develop a model that accounts for unobserved movie attributes in modeling preferences.

Let $C_j = \{i_1, i_2, \ldots, i_{n_j}\}$ represent the index set of the n_j customers who rated movie j. Let r_{ji} represent the rating given by customer i for movie j, where $i \in C_j$. The number of customers that provide ratings for a movie varies, which yields an unbalanced data set. The observations for movie j can be used in specifying a movie-level regression model as follows:

$$r_{ji} = \mathbf{z}'_i\boldsymbol{\beta}_j + e_{ij},\ e_{ij} \sim N(0, \sigma^2), \tag{4}$$

for all $i \in C_j$. The vector $\mathbf{z}_i$ contains customer characteristics for customer i, and $\boldsymbol{\beta}_j$ is a vector of parameters for movie j that represents the movies' appeal structure across customers. The population model that specifies how movies differ in their appeal structures can be written as

$$\boldsymbol{\beta}_j = \mathbf{w}'_j\boldsymbol{\mu} + \boldsymbol{\gamma}_j,\ \boldsymbol{\gamma}_j \sim N(\mathbf{0}, \boldsymbol{\Gamma}), \tag{5}$$

for j = 1 to J. The vector $\mathbf{w}_j$ contains the observed movie characteristics, and $\boldsymbol{\gamma}_j$ represents the unobserved movie effects. The complete model can alternatively be written as

$$r_{ji} = \mathbf{x}'_{ij}\boldsymbol{\mu} + \mathbf{z}'_i\boldsymbol{\gamma}_j + e_{ji},\ e_{ji} \sim N(0, \sigma^2),\ \boldsymbol{\gamma}_j \sim N(\mathbf{0}, \boldsymbol{\Gamma}), \tag{6}$$

for j = 1 to J and for $i \in C_j$. The vector $\mathbf{x}_{ji}$ contains all observed movie attributes and customer characteristics and their interactions. The variance matrix $\boldsymbol{\Gamma}$ provides information about the extent of unobserved heterogeneity in product appeal structures.

Customer and Product Heterogeneity

As is apparent from our previous discussion, recommendation systems operate in contexts that involve many products and customers. It is therefore imperative to account for both customer and product heterogeneity in modeling preferences. We combine both forms of heterogeneity. In the combined model, the rating r_{ij} for customer i can be written as

$$r_{ij} = \mathbf{x}'_{ij}\boldsymbol{\mu} + \mathbf{z}'_i\boldsymbol{\gamma}_i + \mathbf{w}'_j\boldsymbol{\lambda}_i + e_{ji},\ e_{ji} \sim N(0, \sigma^2), \tag{7}$$

$$\boldsymbol{\lambda}_i \sim N(\mathbf{0}, \boldsymbol{\Lambda}),\ \boldsymbol{\gamma}_j \sim N(\mathbf{0}, \boldsymbol{\Gamma}),$$

for i = 1 to I and $j \in M_i$. The vector $\mathbf{x}_{ij}$ contains the movie and customer variables, $\mathbf{z}_i$ is a vector of customer characteristics, and $\mathbf{w}_j$ is a vector of movie attributes. The random effects $\boldsymbol{\lambda}_i$ account for unobserved sources of customer heterogeneity and appear in the model interactively with the observed movie attributes. The random effects $\boldsymbol{\gamma}_j$ account for the unobserved source of heterogeneity in movie appeal structures and interact with the observed customer characteristics. Such a model provides a flexible framework for capturing differences in customer preference structures and movie appeal structures.

We use MCMC methods (Gelfand and Smith 1990) for sampling-based inference. These methods involve sampling parameter estimates from the full conditional distribution of blocks of parameters. In the context of our model, we need to generate random draws for the parameter blocks $\{\boldsymbol{\mu}, \{\boldsymbol{\lambda}_i\}, \{\boldsymbol{\gamma}_j\}, \sigma^2, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}\}$. The Appendix gives the specification of priors and the full conditional distributions.

APPLICATION TO MOVIE RECOMMENDATIONS

In this section, we describe an application of our modeling approach to movie recommendations on the Internet. The choice of the application is guided by two considerations. The first is simply the availability of a large commercial data set. The second is that though several marketing researchers have modeled aggregate movie performance (e.g., Dodds and Holbrook 1988; Eliashberg and Shugan 1997; Jedidi, Krider, and Weinberg 1998; Sawhney and Eliashberg 1996), there is limited work on preferences of individuals.

Data were obtained from an actual recommendation system called EachMovie. This recommendation system was operated by DEC Systems research center for 18 months until September 1997. The public domain data consist of (1) ratings of 75,000 customers for 1628 movies on a six-point scale, (2) movie genre, and (3) user demographics. The original data do not contain expert evaluations from movie critics. We therefore collected expert evaluations for 340 of the movies in the database from various Internet sites and movie directories. We used the ratings data for an arbitrary sample of 2000 customers from those who provided complete demographic information on age and sex. Although it was not a random sample, the choice of 340 movies was arbitrary and was determined more by availability of expert ratings than any other biasing criterion. Overall, our sample data are sparse, consisting of 56,239 of 680,000 (8%) possible ratings. Each person rates between 1 and 235 movies, and the average (median) is 29 (19) movies per person. The data set is also unbalanced, as different customers rate different subsets of movies and different movies are rated by different subsets of people. Overall, the number of ratings per movie ranges from 1 to 1285, and the average (median) is 163 (74) ratings.

Our calibration sample contains 10,344 ratings on 228 movies and 986 customers. We construct four validation samples to reflect the following possible scenarios that can be described in terms of the information that is available on customers and movies in the database. First, a customer can be either an existing one, in which case customer preference data on this customer will be available in the database, or a new one, in which case only demographic information on this customer may be available. Second, a movie can be regarded as an old movie, in which case customer ratings for it are available in the database, or a new one, in which case only genre and expert evaluations are available. Accordingly, we designed our four validation sets to reflect the four possible combinations of customer and movie types (Figure 1).

The first validation sample comprises 2886 observations from the same set of customers and the same set of movies as in the calibration sample. For each person, the holdout movies in this validation sample are different from those the person rates in the calibration sample. Similarly, for each movie, the customers who provide ratings for the movie in this validation sample are different from those in the calibration sample. We label this validation sample old person/old movie. The second validation sample contains ratings from the 986 customers in the calibration sample on

Figure 1
CALIBRATION AND VALIDATION SAMPLES

	Old Movie	New Movie
Old Person	Calibration Old person/old movie validation	Old person/new movie validation
New Person	New person/old movie validation	New person/new movie validation

the remaining 112 holdout movies in the data set. We label this validation sample old person/new movie. The third validation sample contains ratings of the 1014 holdout customers on the 228 movies in the calibration data set. This reflects the situation in which a new customer interacts with the recommendation system. We label this validation sample new person/old movie. The last validation set contains observations on 1014 holdout customers and 112 holdout movies. We label this sample new person/new movie.

Model Specification and Variable Definition

The model with both forms of heterogeneity can be written in terms of four components that include observed variables and two components that include unobserved random effects. The general template for our model is as follows:

$$(8)\quad r_{ij} = \text{Genre}_j + \text{Demographics}_i + \text{Expert Evaluations}_j + \text{Interactions}_{ij} + \text{Customer Heterogeneity}_{ij} + \text{Movie Heterogeneity}_{ij} + e_{ij},$$

where r_{ij} represents customer ratings on a six-point scale from zero to five. We treat these ratings as interval scaled. We now describe each component of our model.

Genre. The genre variables are specified by nine binary indicators that describe whether the movie pertains to one or more of the following genre categories: action, art/foreign, classic, comedy, drama, family, horror, romance, and thriller. A movie can be simultaneously classified into more than one of these genres. The genre effects are included in the model as follows:

$$(9)\quad \text{Genre}_j = \sum_k \mu_k \text{Genre}_{jk},$$

where Genre_{jk} refers to the genre variable k, and μs represent the fixed effects.

Demographics. The demographic variables include age and sex for the customers in the database and are represented in the model as follows:

$$(10)\quad \text{Demographics}_i = \sum_k \mu_k \text{Demographics}_{ik},$$

where Demographics_{jk} refers to the demographic variable k, and μs represent the fixed effects.

Expert evaluations. The expert evaluations are from Roger Ebert (Ebert) of the *Chicago Sun-Times*; James Berardinelli (James) of ReelViews, an online film review site; the Videohound movie directory (Bones); and the Internet Movie Database (IMDB). The variables Ebert, James, and Bones are on a nine-point scale ranging from zero to four in increments of one-half point. The variable IMDB reflects the mean ratings of users on the IMDB and has a range of zero to ten. The expert evaluations effects are included in the model as follows:

$$(11)\quad \text{Expert Evaluations}_j = \sum_l \mu_l \text{Expert}_{jl},$$

where Expert_{jl} refers to the expert variable l, and μs represent the fixed effects as previously.

Interactions. Two types of interactions can be included in the model. First, interactions of the demographics with the different genre and expert variables can be included to capture the impact of observed sources of heterogeneity. Second, interactions between the different movie characteristics can be included to capture the joint impact of different genre and expert variables. Both types of interactions were found to be insignificant across all our specifications. We therefore ignore these for model parsimony.

Customer heterogeneity. We model customer heterogeneity in two ways. First, we allow for a customer-specific random intercept. This captures any idiosyncrasies a customer may exhibit in rating movies in general. Second, we allow for customer-specific interactions between the unobserved customer variables (random effects) and the observed descriptors (genre and expert evaluations) of a movie. The interactions enable us to model how different customers value different movie genres and the opinions of various experts. The customer heterogeneity effects are included in the model as follows:

$$(12)\quad \text{Customer Heterogeneity}_{ij} = \lambda_{il} + \sum_k \lambda_{ik} \text{Genre}_{jk} + \sum_l \lambda_{il} \text{Expert}_{jl},$$

where Genre_{jk} refers to the genre variable k of movie j and Expert_{jl} refers to the expert evaluation l. The λs represent the customer random effects, and the $\boldsymbol{\lambda}_i$ vector is assumed to come from a multivariate normal population distribution $N(\mathbf{0}, \boldsymbol{\Lambda})$.

Product heterogeneity. We model product heterogeneity similarly by allowing (1) a movie-specific random intercept that captures the movie equity and (2) movie-specific interactions between the unobserved movie attributes (random effects) and the observed customer demographics. The interactions enable us to model movie appeal structures, that is, how different aspects of the movie appeal to different customer groups. The movie heterogeneity effects are included in the model as follows:

$$(13)\quad \text{Movie Heterogeneity}_{ij} = \gamma_{jl} + \sum_k \gamma_{jk} \text{Demographics}_{ik},$$

where Demographics_{ik} refers to the demographic variable k of customer i. The γs represent the movie random effects, and the $\boldsymbol{\gamma}_j$ vector is assumed to come from a multivariate normal population distribution $N(\mathbf{0}, \boldsymbol{\Gamma})$.

In Table 1, we summarize the key statistics on the movies and respondent descriptors. For example, Mean 1, computed across all 340 movies and 2000 respondents, indicates that our sample has 16.9% movies and that Roger Ebert gives a mean rating of 2.745 to our movie sample. The average age of our sample respondent is 33 years; 85% of respondents are male. Mean 2 is computed across all 56,239 sample ratings and indicates, for example, that 28.4% of these ratings are for action movies and 83.8% were given by male respondents.

In addition to the complete model, we estimated 11 other restricted models that involve different combinations of the observed effects and heterogeneity specifications. The details of the included and excluded effects are clear from the description of the models in Table 2. The configuration of restricted models enables us to investigate the differential impacts of (1) the two forms of heterogeneity and (2) the different types of movie descriptors in predicting customer preferences. We estimated the models using programs developed in the C language on a Sun Enterprise 4000 machine. The estimation time for 5000 iterations is approximately one and one-half hours, whereas prediction on any given observation takes less than a second.

RESULTS

Model Comparison

We use (1) the marginal likelihood of the data based on the cross-validation predictive density (Gelfand 1996) and (2) the deviance information criterion (DIC) statistic (Spiegelhalter, Best, and Carlin 1998) for model comparison. Both criteria are based on the likelihood of a model and appropriately penalize a model for complexity. Although the marginal likelihood traditionally has been used in the

Table 1
SUMMARY STATISTICS

	Variables	*Mean 1*[a]	*(Standard Deviation 1)*	*Mean 2*[b]	*(Standard Deviation 2)*
Genre variables	Action	.169	(.375)	.284	(.451)
	Art/foreign	.143	(.350)	.080	(.272)
	Classic	.003	(.054)	.007	(.085)
	Comedy	.300	(.459)	.307	(.461)
	Drama	.426	(.495)	.384	(.486)
	Family	.079	(.270)	.108	(.310)
	Horror	.044	(.205)	.074	(.262)
	Romance	.146	(.353)	.152	(.359)
	Thriller	.160	(.367)	.179	(.384)
Expert variables	Ebert	2.745	(.827)	2.888	(.831)
	James	2.653	(.761)	2.738	(.779)
	Bones	2.567	(.651)	2.737	(.656)
	IMDB	7.072	(1.239	6.974	(1.274)
Demographic variables	Sex	.851	(.356)	.838	(.369)
	Age	33.01	(11.28)	31.63	(10.63)

[a]Mean 1 and Standard Deviation 1 are the mean and standard deviation across all 340 movies or all 2000 customers.
[b]Mean 2 and Standard Deviation 2 are the mean and standard deviation across all 56,239 ratings.

Table 2
MODEL COMPARISON STATISTICS

Models			*DIC Statistics*		
Heterogeneity	*Movie Attributes*	*Log-Marginal Likelihood*	*Fit D*	*Complexity pD*	*DIC*
No heterogeneity	Genre only	−18,801	37,589	13	37,602
	Expert only	−18,398	36,788	8	36,796
	Genre and expert	−18,327	36,638	17	36,655
Customer heterogeneity	Genre only	−17,581	34,135	1020	35,155
	Expert only	−17,162	33,429	900	34,329
	Genre and expert	−16,909	32,215	1501	33,716
Movie heterogeneity					
	Genre only	−18,072	35,825	275	36,100
	Expert only	−18,067	35,834	259	36,093
	Genre and expert	−18,066	35,830	260	36,090
Movie and customer Heterogeneity	Genre only	−16,793	32,118	1390	33,508
	Expert only	−16,840	32,502	1146	33,648
	Genre and expert	−16,675	31,488	1717	33,205

Notes: All models include demographic variables.

Bayesian literature, the DIC statistic recently has been suggested as an alternative criterion that is simple to compute in most modeling contexts and can be written as

$$\text{(14)} \qquad DIC = \bar{D}(\boldsymbol{\theta}) + p_D,$$

where $D(\boldsymbol{\theta}) = -2LL(\boldsymbol{\theta})$ is the model deviance and is equal to twice the negative log-likelihood. The vector $\boldsymbol{\theta}$ contains all parameters, including the random effects. The average deviance $\bar{D}$ is computed by taking the average of the deviance over the MCMC draws and is viewed as a measure of model fit. The term p_D penalizes the model for complexity and can be interpreted as the effective number of parameters in the model. It is computed as $p_D = \bar{D}(\boldsymbol{\theta}) - D(\bar{\boldsymbol{\theta}})$, where $D(\bar{\boldsymbol{\theta}})$ is the deviance calculated using the mean of the parameters $\bar{\boldsymbol{\theta}}$ obtained from the MCMC draws. The model with the lowest DIC is considered the best model.

In Table 2, we report the log-marginal likelihoods and the DIC statistics for all our models. We show that the complete model (last row) outperforms all other models on both model comparison criteria. The DIC statistics in the last row indicate that though this model is the most complex (p_D = 1717), it is also the best fitting (Fit = 31,488), and therefore it outperforms the other models on the DIC statistic. In contrast, in the first row of Table 2, we show that the model that does not capture any source of unobserved heterogeneity and includes only genre variables for describing the movies performs the worst.

A comparison of the different classes of models shows that the first set of models that do not allow for unobserved heterogeneity has the least support, whereas the models that include both sources of unobserved heterogeneity (the last set) have the greatest empirical support on both model comparison criteria. A comparison of the second set of models with the third set shows that accounting for customer heterogeneity is more important than accounting for movie heterogeneity. The improvements in log-marginal likelihood are greater when customer heterogeneity is added (especially in the impact of expert variables) to the first set of models than when movie heterogeneity is included as in the third set. The DIC statistics also imply that the more complex models with customer heterogeneity outperform models with movie heterogeneity alone.

Parameters Estimates

In Table 3, we report the posterior mean estimates for the fixed effects $\boldsymbol{\mu}$ and the standard deviations (i.e., the square root of the diagonal elements of $\boldsymbol{\Lambda}$ of $\boldsymbol{\Gamma}$) of the customer and movie random effects. Note that a variable can influence preferences either directly through the fixed effects or indirectly by interacting with the random effects. The first row of Table 3 shows that the standard deviation of the customer-specific random effect associated with the intercept is 1.647. This implies that customers differ in their use of the rating scale. Similarly, the movie equity differs across the movies in the sample. This is evident from the fifth column in Table 3, which shows that the standard deviation of the movie-specific random effect associated with the intercept is .515. Most genre variables show insignificant fixed effects. On average, people like action and thriller movies and dislike horror movies. The standard deviations of customer-specific random effects pertaining to the genre variables, however, are large for all genre variables and unambiguously indicate that customers differ in their preference structures. Thus, accounting for differences in the preference structures across customers is important for our application.

The fixed effects for James, Bones, and IMDB are positive and significant and imply that the expert evaluations are, in general, positively associated with the ratings in the database. The random effects pertaining to the expert variables vary significantly across the users, as is evident from the significant magnitudes of their standard deviations across customers. This implies that the association of the ratings with the expert evaluations varies across customers. Thus, accounting for expert evaluations is crucial in this application. Finally, the fixed effect for sex is insignificant, whereas the coefficient for age is significant. The demographics are an important component of the model, as the standard deviations of the associated random effects across the movies reveal differences in movie appeal structures. Thus, it appears that movie appeal differs across demographic groups, but this difference is based on unobserved movie attributes.

Predictive Ability

In Table 4, we report the root mean square errors (RMSEs) in prediction for all the models on the calibration data and on each of the four validation data sets. In general, the RMSE statistics decrease in magnitude from the top to the bottom of the table. Table 4 also shows that models that include customer heterogeneity have better fit than models that include movie heterogeneity alone. Comparing the RMSE statistics for the proposed model (last row) with those obtained from the restricted models, we see that the proposed model outperforms the other models on almost all the data sets.

How well does the model compare with actual recommendation systems on the market? Unfortunately, commercial vendors are loath to share proprietary implementations, and any implementation of algorithms we execute is open to criticism by these companies. Fortunately, a recent report (Breese, Heckerman, and Kadie 1998), written by researchers at Microsoft, Firefly's parent company, gives us some bases for comparing our method with implementations of four collaborative filtering algorithms for the same database of movies used in the present article. These four methods are (1) collaborative filtering, (2) Bayesian networks, (3) Bayesian clustering (mixture/latent-class models), and (4) vector similarity. Using all but one observation for model estimation, Breese, Heckerman, and Kadie (1998) find that the excluded movie is, on average, approximately one rating point away from its true value. Specifically, if mean absolute deviation (MAD) = (Σ|predicted rating – true rating|/number of observations) is the MAD of the predicted and true ratings over the holdout movies, its value is .994 for collaborative filtering, 1.103 for Bayesian clustering, and 1.006 for Bayesian networks. Vector similarity, with a value of MAD = 2.136, does the worst.

To assess the predictive accuracy of our models on the original scale, we transform the continuous predictions to a zero to five scale. We use an optimal set of thresholds (much as in an ordinal probit specification) to transform the predictions. We obtain these thresholds by minimizing the MADs on the calibration data through a grid search. The proposed model, estimated using all but one holdout movie per person, gives MAD = .899 for the holdout set.[3]

[3]Replication on another holdout set yields a MAD = .79.

Moreover, whereas Breese, Heckerman, and Kadie's (1998) test uses an average of 46.5 movies per person (median = 36), we use an average of 17 movies per person (median = 10). Even if we use an average of 12.82 movies per person (median = 8), our model produces MAD = .905 on a holdout set comprising on average 5.33 movies per person (median = 3). However, before we conclude that our method is superior to these others, it is worth examining the performance of an aggregate regression—an uncustomized recommendation system—that uses expert ratings and genre as the independent variables. The resulting MAD = 1.094 suggests that none of the models—including ours—does particularly well compared with this baseline model.[4]

Why does regression do so well? Not because it is a better model, but because MAD is not very informative about how a model fails. To see this, compare the row/column marginals of the incidence matrices in Tables 5 and 6. Note that (1) 54.19% of cases are true 3/4 ratings, (2) our model predicts 3/4 ratings in 66.28% of cases, and (3) regression predicts 3/4 in an overwhelming 90.3% of cases. That is, the aggregate regression makes virtually identical predictions for almost all people and movies. It is not a good method for making recommendations, but it gets close on average, at least in this data set.

A better approach is to examine the full information—the 6 × 6 matrix of actual versus predicted ratings. Tables 5 and 6 display these for the proposed method and regression for the comparison reported previously, which is restricted to old person/old movie. Table 5 can be interpreted as follows: If we predict a rating of 0 for a movie that is not seen by a person, the odds that a person who sees it will rate it a 0 is 65%, and so forth. That is, instead of making a point prediction, we can present to users the odds that if they see a new movie they will subsequently give it a certain rating. Three immediate observations can be made from an examination of Table 5. First, the diagonal numbers (perfect pre-

[4]The regression also uses a holdout set with an average of 5.33 movies per person. We caution that a direct comparison of results is not possible, because the calibration data set we use differs from the exact data set used in Breese, Heckerman, and Kadie's (1998) study.

Table 3

PARAMETER ESTIMATES FOR THE COMPLETE MODEL

	Variables	*Fixed Effects* μ^a	*Standard Deviation Across Customers*	*Standard Deviation Across Movies*
	Intercept	.325 (.371)	1.647	.515
Genre variables	Action	.341* (.131)	.407	
	Art/foreign	.053 (.148)	.573	
	Classic	–.085 (.571)	.318	
	Comedy	.210 (.127)	.300	
	Drama	.124 (.117)	.257	
	Family	.048 (.168)	.321	
	Horror	–.411* (.226)	.347	
	Romance	–.084 (.137)	.301	
	Thriller	.319* (.162)	.292	
Expert variables	Ebert	.085 (.062)	.127	
	James	.231* (.089)	.210	
	Bones	.266* (.088)	.229	
	IMDB	.125* (.048)	.087	
Demographic variables	Sex	.018 (.074)		.155
	Age	.006* (.003)		.024
	$\sigma 2$	1.229* (.021)		

[a]Standard deviations across MCMC iterations are in parentheses.
*Significant at the .05 level.

dictions) are not impressive. Second, the percentage of perfect matches decreases toward the center of the scale. Third and most important, the greatest proportion of errors are nearest neighbors. In other words, if we make a forecast of 5, then 86% of the true ratings are 4 or 5. These nearest neighbor percentages can be summarized as in Table 7.

From these numbers, we would be most confident in making point predictions within a rating point at the top end of the scale. This is a fortunate coincidence, which may or may not carry over to other applications. Our model will never give recommendations for movies with low ratings. And for movies with high predicted ratings (4 or 5), it will do quite well. A recommendation system built on the basis of these results should restrict recommendations to movies with ratings of 4 or 5. If the predicted rating is 4, the actual rating is no lower than 4 in 68% of cases and no lower than 3 in 90% of cases: It is a good surprise 26% of the time and a mild disappointment 22% of the time. Translated to raw numbers (see Table 5), this means that of the 1022 recommendations we make with a rating of 4, in 427 cases, people who see these movies will come away with the exact same judgment about the movie as we predict; there will be 266 good surprises, and 224 will be mildly disappointed (we use the terms "happier" and "slightly" colloquially, for there is no saying how unhappy a slightly disappointed person might be). Among the remaining, 43 will probably not want to use the recommendation system again. Similarly, if people see the 325 recommendations for which we predict a rating of 5, there will be complete agreement with our assessment in 165 cases, another 116 cases of mild disappointment, and 44 cases of reactions that range from disappointment to sheer exasperation.

How disappointed will people be if they are not told of a movie they would have liked? The answer depends on our criteria for not recommending. If we never recommend movies for which we predict a rating of 0, then .5% of people (11 of 2886 people in our sample) will not see a movie they would rate 4 or 5 were they to see it, because we failed

Table 4

ROOT MEAN SQUARE ERRORS

		Calibration Sample	*Old Person/ Old Movie*	*Old Person/ New Movie*	*New Person/ Old Movie*	*New Person/ New Movie*
No heterogeneity	Genre only	1.488	1.504	1.544	1.421	1.500
	Expert only	1.431	1.453	1.460	1.376	1.444
	Genre and expert	1.420	1.442	1.476	1.370	1.458
Movie heterogeneity	Genre only	1.349	1.442	1.542	1.349	1.468
	Expert only	1.350	1.417	1.473	1.348	1.460
	Genre and expert	1.350	1.418	1.488	1.349	1.478
Customer heterogeneity	Genre only	1.196	1.313	1.307	1.414	1.516
	Expert only	1.163	1.278	1.299	1.369	1.438
	Genre and expert	1.061	1.241	1.306	1.361	1.456
Movie and customer heterogeneity	Genre only	1.063	1.233	1.367	1.339	1.483
	Expert only	1.192	1.287	1.296	1.383	1.432
	Genre and expert	1.012	1.216	1.295	1.337	1.446

Table 5

PROPOSED MODEL: VALIDATION SAMPLE OLD PERSON/OLD MOVIE INCIDENCE MATRIX

	Predicted						
Actual	*0*	*1*	*2*	*3*	*4*	*5*	*Total*
0	130* (64.68)**	67 (37.22)	56 (20.97)	58 (6.51)	24 (2.35)	3 (.92)	338 (11.71)
1	23 (11.44)	21 (11.67)	26 (9.74)	55 (6.17)	19 (1.86)	1 (.31)	145 (5.02)
2	19 (9.45)	31 (17.22)	44 (16.48)	123 (13.80)	62 (6.07)	10 (3.08)	289 (10.01)
3	18 (8.96)	42 (23.33)	82 (30.71)	296 (33.22)	224 (21.92)	30 (9.23)	692 (23.98)
4	7 (3.48)	13 (7.22)	43 (16.10)	266 (29.85)	427 (41.78)	116 (35.69)	872 (30.21)
5	4 (1.99)	6 (3.30)	16 (5.99)	93 (10.44)	266 (26.03)	165 (50.77)	550 (19.26)
Total percentage	201 (6.96)	180 (6.24)	267 (9.25)	891 (30.87)	1022 (35.41)	325 (11.26)	2886 (100)

*Cell frequency.

**Column percentage.

Table 6
REGRESSION: VALIDATION SAMPLE OLD PERSON/OLD MOVIE INCIDENCE MATRIX

	Predicted						
Actual	*0*	*1*	*2*	*3*	*4*	*5*	*Total*
0	0* (0)**	37 (38.54)	29 (31.52)	147 (14.89)	124 (7.66)	1 (1.11)	338 (11.71)
1	1 (50)	13 (13.54)	11 (11.96)	68 (6.89)	49 (3.03)	3 (3.33)	145 (5.02)
2	1 (50)	15 (15.63)	12 (13.04)	131 (13.27)	126 (7.78)	4 (4.44)	289 (10.01)
3	0 (0)	17 (17.71)	28 (30.43)	288 (29.18)	346 (21.37)	13 (14.44)	692 (23.98)
4	0 (0)	13 (13.54)	9 (9.78)	247 (25.03)	568 (35.08)	35 (38.89)	872 (30.21)
5	0 (0)	1 (1.04)	3 (3.26)	106 (10.74)	406 (25.08)	34 (37.78)	550 (19.06)
Total percentage	2 (.07)	96 (3.33)	92 (3.19)	987 (34.20)	1619 (56.10)	90 (3.12)	2886 (100)

*Cell frequency.
**Column percentage.

Table 7
NEAREST NEIGHBOR PERCENTAGES

Predicted	*Nearest Neighbors*	*Percent Actual Within One Point of Prediction*
5	4	86
4	3, 5	90
3	2, 4	77
2	1, 3	57
1	0, 2	66
0	1	76

to tell them about it. The 2 and 3 predictions are murky grounds. Many movies are good to miss, and many are worth seeing. In summary, if we consider only exact matches, our prediction model is far from perfect. But for making recommendations, it is not so bad at all.

A priori, the quality of recommendations should decrease as the available information about a person or movie decreases. This is indeed the case for MAD, which ranges from .905 for old movies/old people to 1.122 for new movies/new people; the values for the two other conditions are .971 (old person/new movie) and 1.008 (new person/old movie). Nevertheless, even in the worst case, the deterioration in the MAD values is not terrible, as it changes from approximately 1 point for all the previous methods to 1.12 in the worst case (new person/new movie). However, the full distribution of predictions must be examined to assess how well the method does in each case. In Tables 8 to 10, we give the complete distributions. The odds of true rating when ratings of 5 (which correspond to movies that might be recommended) are predicted do not differ markedly from each other, except for new person/new movie. But for old person/old movie, the percentage of movies predicted to have 5 ratings decreases from 11.26% to 7.17%, whereas the percentage of movies with true ratings of 5 drops from 19.26% to 17.14%. In other words, the model becomes more conservative in predicting a 5 rating from old to new movies. This conservative pattern of predictions becomes even more pronounced as less information is used to make predictions. Thus, in the worst case (new person/new movie), the model predicts ratings of 5 in 3.95% of cases, whereas 17.76% of movies have a true rating of 5, a number close to the 3.12% for the aggregate regression in Table 6. A conservative pattern of predictions is also observed at the other end of the scale (i.e., for 0 and 1 ratings).

A similar analysis when 4 ratings are predicted suggests a more noticeable decline in the performance of the model: The proportion of true ratings with a value of 3 increases significantly as less information becomes available. Still, if the odds of a 4 prediction being a true 4 or 5 rating are considered, the worst case (new movie/new person) odds of .6 are not bad compared with the odds of .678 in the best case (old movie/old person).

In summary, compared with collaborative filtering, the proposed model is substantially better for two reasons. First, it does better when both collaborative filtering and our method can be used to make recommendations (i.e., old person/old movie). Second, it can make recommendations when collaborative filtering cannot be used. As might be expected, the less information about a person or a movie, the less accurate are the predictions.

As long as making recommendations is the goal, the present model performs well, because this entails making accurate predictions of movies with high (4 or 5) ratings, provided there are enough movies for which these ratings are predicted. But if responding to queries (Should I see this movie or not?) is the goal, even in the best case (old movie/old person), a predicted rating of 2 or 3 has unacceptably high odds of being some other true rating. This is an area in which model improvement is especially desirable. That said, we believe that it is better for any recommendation system not just to make a point prediction but also to predict odds for true ratings in the manner discussed here, together with the count (number of observations) on which the odds are based. This, in the end, is much better for mak-

Table 8

PROPOSED MODEL: VALIDATION SAMPLE OLD PERSON/NEW MOVIE INCIDENCE MATRIX

	Predicted						
Actual	*0*	*1*	*2*	*3*	*4*	*5*	*Total*
0	549* (62.10)**	273 (32.85)	324 (20.02)	467 (7.55)	136 (2.58)	26 (2.27)	1775 (11.14)
1	84 (9.50)	113 (13.60)	177 (10.94)	417 (6.74)	138 (2.62)	29 (2.54)	958 (6.01)
2	91 (10.29)	160 (19.25)	298 (18.42)	806 (13.04)	329 (6.24)	53 (4.64)	1737 (10.90)
3	84 (9.50)	155 (18.65)	455 (28.12)	1912 (30.92)	1119 (21.21)	106 (9.27)	3831 (24.04)
4	56 (6.33)	95 (11.43)	271 (16.75)	1895 (30.65)	2220 (42.08)	366 (32.02)	4903 (30.77)
5	20 (2.26)	35 (4.21)	93 (5.75)	686 (11.09)	1334 (25.28)	563 (49.26)	2731 (17.14)
Total percentage	884 (5.55)	831 (5.21)	1618 (10.15)	6183 (38.80)	5276 (33.11)	1143 (7.17)	15,935 (100)

*Cell frequency.
**Column percentage.

Table 9

PROPOSED MODEL: VALIDATION SAMPLE NEW PERSON/OLD MOVIE INCIDENCE MATRIX

	Predicted						
Actual	*0*	*1*	*2*	*3*	*4*	*5*	*Total*
0	6* (54.55)**	132 (36.97)	158 (20.08)	365 (8.76)	274 (4.92)	14 (1.68)	949 (8.09)
1	2 (18.18)	31 (8.68)	95 (12.07)	254 (6.10)	137 (2.46)	9 (1.08)	528 (4.50)
2	1 (9.09)	55 (15.41)	126 (16.01)	510 (12.24)	362 (6.50)	20 (2.40)	1074 (9.16)
3	1 (9.09)	69 (19.33)	191 (24.27)	1108 (26.60)	1095 (19.65)	99 (11.86)	2563 (21.86)
4	0 (.00)	54 (15.13)	165 (20.97)	1364 (32.75)	2090 (37.51)	266 (31.86)	3939 (33.59)
5	1 (9.09)	16 (4.48)	52 (6.61)	564 (13.54)	1614 (28.97)	427 (51.14)	2674 (22.80)
Total percentage	11 (.00)	357 (3.00)	787 (6.70)	4165 (35.5)	5572 (47.5)	835 (7.1)	11,727 (100)

*Cell frequency.
**Column percentage.

ing informed choices than a forced prediction, which in situations involving risky decisions, can lead to bad decisions by a person who places undue trust in point predictions.

SUMMARY AND CONCLUSIONS

The simple, flexible models we describe can be generalized to incorporate revealed preferences on the basis of explicit or implicit data. Methodologically, it seems useful to consider the predictive improvements possible if nonlinearity is incorporated by means of radial basis function neural networks, multiplayer perceptrons, or other methods. Alternative methods, such as latent-class models, may also be worth investigating. We developed procedures for ratings data. Extensions for ordinal or binary data can be handled within the Bayesian framework with data augmentation methods (Albert and Chib 1993). Similarly, data augmentation can be used to handle censoring that may arise because rated movies have a higher or lower rating compared with unrated movies. Finally, recommendation systems represent just one type of "agent." A variety of other information agents—for example, negotiation agents, matchmaking agents, and agents designed to participate in auctions—are directly relevant for marketers. The approaches and methodologies that have evolved in the marketing literature to explain customer preferences and other aspects of consumer behavior can likely be used in each of these domains. The new applications of information agents will also require advances in data collection and analysis procedures; marketing researchers are eminently poised to contribute significantly in these areas.

APPENDIX: PRIOR DISTRIBUTIONS AND FULL CONDITIONAL DISTRIBUTIONS

Priors

The unknown parameters for the model are $\beta = \{\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \sigma\}$ In this article, we specify the prior distribution over β as a product of independent priors. We use proper but diffuse priors over all model parameters. The prior for $\boldsymbol{\mu}$ is multivariate normal $N(\boldsymbol{\eta}, \mathbf{C})$ The covariance matrix $\mathbf{C}$ is diagonal, and large values for the variances reflect uncertainty. We use $\boldsymbol{\eta} = \mathbf{0}$ and $\mathbf{C} = 1000\mathrm{I}$, where I is the identity matrix. The precision matrix $\boldsymbol{\Lambda}^{-1}$ associated with the population distribution $\boldsymbol{\lambda}_i \sim N(\mathbf{0}, \boldsymbol{\Lambda})$ is a $(m + 1) \times (m + 1)$ positive definite matrix, where m is the number of movie descriptors. Similarly, the precision matrix $\boldsymbol{\Gamma}^{-1}$ associated with the population distribution $\boldsymbol{\gamma}_j \sim N(\mathbf{0}, \boldsymbol{\Gamma})$ is a $(p + 1) \times (p + 1)$ positive definite matrix, where p is the number of customer characteristics. We assume Wishart priors: $W[\iota, (\iota\mathbf{L})^{-1}]$ for the precision matrix $\boldsymbol{\Lambda}^{-1}$, and $W[\jmath, (\jmath\mathbf{G})^{-1}]$ for the precision matrix $\boldsymbol{\Gamma}^{-1}$. The matrices $\mathbf{L}$ and $\mathbf{G}$ can be considered the expected prior variances of the $\boldsymbol{\lambda}_i$s and $\boldsymbol{\gamma}_j$s, respectively. Smaller values for ι and $\jmath$ correspond to more diffuse prior distributions. We set $\iota = 16$, $\jmath = 5$, $\mathbf{G} = \mathrm{diag}(.001)$, and $\mathbf{L} = \mathrm{diag}(.001)$. The prior for the error variance σ^2 is chosen to be inverse gamma IG(a, b), where a = 3 and b = 10.

Full Conditional Distributions for the Full Model

The parameter $\boldsymbol{\mu}$ can be generated from the multivariate normal full conditional distribution given by

$$\text{(A1)} \qquad p(\boldsymbol{\mu} | \{r_{ij}\}, \{\boldsymbol{\lambda}_i\}, \{\boldsymbol{\gamma}_j\}, \sigma^2) \sim N(\hat{\boldsymbol{\mu}}, \mathbf{V}_\mu),$$

where $\mathbf{V}_\mu^{-1} = C^{-1} + \sigma^{-2}\mathbf{X}'\mathbf{X}$, and $\hat{\boldsymbol{\mu}} = \mathbf{V}_\mu(\sigma^{-2}\mathbf{X}'\tilde{\mathbf{r}} + C^{-1}\boldsymbol{\eta})$. The matrix $\mathbf{X}$ is obtained by stacking row by row all the row vectors $\mathbf{x}'_{ij}$. The vector $\tilde{\mathbf{r}}$ is obtained by stacking all the elements $\tilde{r}_{ij} = r_{ij} - \mathbf{z}'_i\boldsymbol{\gamma}_j - \mathbf{w}'_j\boldsymbol{\lambda}_i$, for all the person–movie pairs.

The full conditional distribution of the error variance σ^2 is inverse gamma and is given by

$$\text{(A2)} \qquad p(\sigma^2 | \{r_{ij}\}, \{\boldsymbol{\lambda}_i\}, \{\boldsymbol{\gamma}_j\}, \boldsymbol{\mu})$$

$$\sim IG\left\{\frac{N}{2} + a\left[\frac{1}{2}(\tilde{\mathbf{r}} - \mathbf{X}\boldsymbol{\mu})'(\tilde{\mathbf{r}} - \mathbf{X}\boldsymbol{\mu}) + b^{-1}\right]^{-1}\right\}$$

The customer random effects $\boldsymbol{\lambda}_i$ can be generated from the multivariate normal full conditional distribution given by

$$\text{(A3)} \qquad p(\boldsymbol{\lambda}_i | \{r_{ij}\}, \boldsymbol{\mu}, \{\boldsymbol{\gamma}_j\}, \sigma^2, \boldsymbol{\Lambda}) \sim N(\hat{\boldsymbol{\lambda}}_i, V_i),$$

where $\mathbf{V}_i^{-1} = \boldsymbol{\Lambda}^{-1} + \sigma^{-2}\mathbf{W}'_i\mathbf{W}_i$ and $\hat{\boldsymbol{\lambda}}_i = \mathbf{V}_i(\sigma^{-2}\mathbf{W}'_i\tilde{\mathbf{r}}_i)$. The matrix $\mathbf{W}_i$ is obtained by stacking row by row all the row vectors $\mathbf{w}'_j$ for j belonging to the index set of customer i's movies, M_i. The vector $\tilde{\mathbf{r}}_i$ is obtained by stacking the elements $\tilde{r}_{ij\lambda} = r_{ij} - x'_{ij}\boldsymbol{\mu} - \mathbf{z}'_i\boldsymbol{\gamma}_j$, for all the movies $j \in M_i$ of customer i.

The movie random effects $\boldsymbol{\gamma}_j$ can be generated from the multivariate normal full conditional distribution given by

$$\text{(A4)} \qquad p(\boldsymbol{\gamma}_j | \{r_{ij}\}, \boldsymbol{\mu}, \{\boldsymbol{\lambda}_i\}, \sigma^2, \boldsymbol{\Gamma}) \sim N(\hat{\boldsymbol{\gamma}}_j, \mathbf{V}_j),$$

where $\mathbf{V}_j^{-1} = \boldsymbol{\Gamma}^{-1} + \sigma^{-2}\mathbf{Z}'_j\mathbf{Z}_j$ and $\hat{\boldsymbol{\gamma}}_j = \mathbf{V}_j(\sigma^{-2}\mathbf{Z}'_j\tilde{\mathbf{r}}_j)$. The matrix $\mathbf{Z}_j$ is obtained by stacking all the row vectors $\mathbf{z}'_i$ for i in C_i, the index set of movie j's customers. The vector $\tilde{\mathbf{r}}_j$ is obtained by stacking the elements $\tilde{r}_{ij\lambda} = r_{ij} - x'_{ij}\boldsymbol{\mu} - \mathbf{w}'_j\boldsymbol{\lambda}_i$ for all the customers $i \in C_j$ of movie j.

The full conditional distribution of the precision matrix $\boldsymbol{\Lambda}^{-1}$ of the unobserved customer characteristics is Wishart and is given by

$$\text{(A5)} \qquad p\left(\Lambda^{-1} \mid \{\boldsymbol{\lambda}_i\}\right) \sim W\left[\left(\sum_{i=1}^{I} \lambda_i \lambda_i' + \iota L\right)^{-1}, \iota + I\right].$$

Table 10

PROPOSED MODEL: VALIDATION SAMPLE NEW PERSON/NEW MOVIE INCIDENCE MATRIX

	Predicted						
Actual	*0*	*1*	*2*	*3*	*4*	*5*	*Total*
0	2*	73	137	947	314	48	1521
	(100)**	(45.63)	(14.50)	(10.85)	(6.29)	(7.87)	(9.86)
1	0	16	94	522	191	24	847
	(.00)	(10.00)	(9.95)	(5.98)	(3.83)	(3.93)	(5.49)
2	0	15	153	1101	399	44	1712
	(.00)	(9.38)	(16.19)	(12.62)	(8.00)	(7.21)	(11.09)
3	0	33	255	2257	1088	84	3717
	(.00)	(20.63)	(26.98)	(25.87)	(21.81)	(13.77)	(24.09)
4	0	15	233	2742	1755	149	4894
	(.00)	(9.38)	(24.66)	(31.43)	(35.18)	(24.43)	(31.72)
5	0	8	73	1156	1242	261	2740
	(.00)	(5.00)	(7.72)	(13.25)	(24.89)	(42.79)	(17.76)
Total	2	160	946	8725	4989	610	15,431
percentage	(.01)	(1.04)	(6.12)	(56.54)	(32.33)	(3.95)	(100)

*Cell frequency.

**Column percentage.

The full conditional distribution of the precision matrix Γ^{-1} of the unobserved movie characteristics is also Wishart and is given by

$$\text{(A6)} \quad p\left(\Gamma^{-1} \mid \{\gamma_j\}\right) \sim W\left[\left(\sum_{j=1}^{J} \gamma_j \gamma'_j + jG\right)^{-1}, \ j + J\right]$$

REFERENCES

Albert, J. and S. Chib (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88 (June), 669–79.

Allenby, G. and J. Ginter (1995), "Using Extremes to Design Products and Segment Markets," *Journal of Marketing Research*, 32 (November), 392–403.

Balabanović, Marko and Yoav Shoham (1997), "Fab: Content Based Collaborative Recommendation," *Communications of the Association for Computing Machinery*, 40 (3), 66–72.

Breese, John S., David Heckerman, and Carl Kadie (1998), "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," technical report MSR-TR-98-12, Microsoft Research, (October).

Chien, Y.-H. and Edward I. George (1999), "A Bayesian Model for Collaborative Filtering," technical report, Statistics Department, University of Texas at Austin.

Dodds, John C. and Morris B. Holbrook (1988), "What's an Oscar Worth? An Empirical Estimation of the Effect of Nominations and Awards on Movie Distribution and Revenues," in *Current Research in Film: Audiences, Economics and the Law*, Vol. 4, B.A. Austin, ed. Norwood, NJ: Ablex Publishing Co., 72–88.

Eliashberg, Jehoshua and Steven M. Shugan (1997), "Film Critics: Influencers or Predictors?" *Journal of Marketing*, 61 (2), 68–78.

Foner, Leonard N. (1993), "What Is an Agent, Anyway? A Sociological Case Study" working paper, MIT Media Labs.

Gelfand, A.E. (1996), "Model Determination Using Sampling-Based Methods," in *Markov Chain Monte Carlo in Practice*, W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, eds. London, Chapman & Hall, 145–58.

——— and A.F.M. Smith (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85 (June), 398–409.

Gershoff, Andrew and Patricia M. West (1998), "Using a Community of Knowledge to Build Intelligent Agents," *Marketing Letters*, 9 (February), 79–91.

Goldberg, David, David Nichols, Brian M. Oki, and Douglas Terry (1992), "Using Collaborative Filtering to Weave an Information Tapestry," *Communications of the Association for Computing Machinery*, 35 (12), 61–70.

Heckerman, David (1996), "A Tutorial on Learning with Bayesian Networks," Technical Report, MSR-TR-95-06, Microsoft Corporation.

Jedidi, Kamel, R.E. Krider, and CB. Weinberg (1998), "Clustering at the Movies," *Marketing Letters*, 9 (4), 393–405.

Kay, Alan (1984), "Computer Software," *Scientific American*, 251 (September), 53–59.

Levy, Steven (1993), "Let Your Agents Do the Walking," *MacWorld*, (May), 42.

Negroponte, Nicholas (1970), *The Architecture Machine*. Boston: MIT Press.

Rossi, P., R. McCulloch, and G. Allenby (1996), "On the Value of Household Purchase History Information in Target Marketing," *Marketing Science*, 15 (4), 321–40.

Salton, Gerard and Christopher Buckley (1988), "Term Weighting Approaches in Automatic Text Retrieval," *Information Processing and Management*, 24 (5), 513–23.

Sarwar, Badrul M., Joseph A. Konstan, Al Borchers, John Herlocker, Brad Miller, and John Riedl (1998), "Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System," in *Proceedings of the ACM 1998 Conference on Computer Supported Collaborative Work*. New York: ACM Press, 345.

Sawhney, Mohanbir S. and Jehoshua Eliashberg (1996), "A Parsimonious Model for Forecasting Gross Box Office Revenues of Motion Pictures," *Marketing Science*, 15 (2), 113–31.

Shardanand, Upendra and Pattie Maes (1995), "Social Information Filtering: Algorithms for Automating 'Word of Mouth,'" in *Proceedings of the ACM CHI '95 Conference on Human Factors in Computing Systems*. New York: ACM Press, 210–17.

Spiegelhalter, David J., N.G. Best, and B.P. Carlin (1998) "Bayesian Deviance, the Effective Number of Parameters, and the Comparison of Arbitrarily Complex Models," technical report, Division of Biostatistics, University of Minnesota.

West, P.M., D. Ariely, S. Bellman, E. Bradlow, J. Huber, E. Johnson, B. Kahn, J. Little, and D. Schkade (1999), "Agents to the Rescue?" *Marketing Letters*, 10 (August), 285–300.