

Recommender Systems and their Impact on Sales Diversity

Daniel Fleder and Kartik Hosanagar
The Wharton School, University of Pennsylvania
{dfleder,kartikh}@wharton.upenn.edu

ABSTRACT

This paper examines the effect of recommender systems on the diversity of sales. Two anecdotal views exist about such effects. Some believe recommenders help consumers discover new products and thus increase sales diversity. Others believe recommenders only reinforce the popularity of already popular products. This paper is a first attempt to reconcile these seemingly incompatible views. We explore the question in two ways. First, modeling recommender systems analytically allows us to explore their path dependent effects. Second, turning to simulation, we increase the realism of our results by combining choice models with actual implementations of recommender systems. We arrive at three main results. One, some common recommenders lead to a net reduction in average sales diversity. Two, there exists path dependence, and in individual instances the same recommender can either increase or decrease diversity. Three, we show how basic design choices affect the outcome.

Categories and Subject Descriptors

H.4.0 [Information Systems Applications]

General Terms

Algorithms, Management, Measurement, Design, Economics.

Keywords

Business, economics, electronic commerce, recommender systems, collaborative filtering, path dependence, simulation, long tail, concentration, diversity.

1. INTRODUCTION

Media has historically been a “blockbuster” industry. Of the many products available, sales concentrate among a small group of hits. In recent years, such concentration has begun to decrease. Online firms have been able to offer more products, and consumers have taken to such expanded offerings (Anderson, 2004; Brynjolfsson et al. 2006; Clemons et al., 2006). As firms offer more products, consumers need a means to find and evaluate these many

alternatives. Recommender systems have long been considered one solution this problem (Pham & Healey, 2005). Firms too have an incentive to provide these systems because they can increase profits (Schafer et al., 1999), which explains their popularity at major online firms such as Amazon, Netflix, and Apple’s iTunes Store. It is a popular belief that recommenders facilitate diversity, but we present an argument why this might not be so. Will recommenders lead us to all become, for example, viewers of niche, independent movies and music? Or, as we investigate, might they actually reinforce the existing blockbuster nature of media? Holding supply-side offerings fixed, we ask whether recommenders do indeed make our media consumption more diverse.

Anecdotally, two views exist about such effects. As stated, some believe recommender systems help consumers discover new products and thus increase sales diversity (e.g., Brynjolfsson, 2006). A small number of others believe recommenders might only reinforce the popularity of already popular products and thus reduce sales diversity (e.g., Mooney & Roy, 2000). This paper is a first attempt to reconcile these seemingly incompatible views.

We explore this question in two ways. First, modeling recommender systems analytically allows us to explore their path dependent effects. Second, turning to simulation, we increase the realism of our results by combining choice models with actual implementations of recommender systems. We arrive at three main results. One, some common recommenders lead to a net reduction in average sales diversity. Two, there exists path dependence, and in individual instances the same recommender can either increase or decrease sales diversity. Three, we show how basic design choices affect the outcome.

2. PRIOR WORK

The design of recommender systems is an active research area. Reviews are provided in Breese et al. (1998) and Adomavicius & Tuzhilin (2005). For commercial applications, authors such as Schafer et al. (1999), Sarwar et al. (2001), and Linden et al. (2003) describe implementations at firms such as Amazon.com and CDNOW. While there is now a large body of work on building these systems, we know much less about how they affect behavior.

At the consumer level, Senecal et al. (2004) show experimentally that recommendations do influence choice. They find that often online systems can be even more influential than human recommendations. Cooke et al. (2002) examine how the likelihood of purchasing a recommended product depends on the information provided, context, and familiarity. From an economic perspective, Pennock et al. (2000) consider the social implications

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

EC’07, June 11-15, 2007, San Diego, California, USA.

Copyright 2007 ACM 978-1-59593-653-0/07/0006...\$5.00.

of collaborative filters. While their focus is to determine whether recommendations satisfy properties from social choice theory (e.g., the weak Pareto property, independence of irrelevant alternatives), we investigate their effects on consumer behavior.

While the above studies ask how recommenders affect single individuals, our interest is the aggregate effect that recommenders have on society and markets. In particular, we are concerned with changes in sales diversity, which is an aggregate measure. To the best of our knowledge, there have not been formal studies isolating such effects, although the topic has received mention from several researchers. Mooney and Roy (2000) suggest collaborative filters may perpetuate concentration, but it is an in-passing comment without formal study. The “cold-start” problem of recommendations is that unpurchased/unrated items cannot be recommended by collaborative filters (Schein et al., 2002); this also suggests recommenders might perpetuate concentration, although it is not the authors’ specific focus. Brynjolfsson et al. (2006) study the increased diversity of online versus offline sales. They discuss several demand-side drivers but do not isolate the specific effect of recommenders. The information cascades literature has looked at aggregate effects of observational learning (Bikhchandani, 1998). For recommenders that use data on prior outcomes, this work has a similar theme. Drawing parallels is difficult, however, because of the complexity of the recommender system algorithm and recommender-consumer interaction.

This prior work reveals four themes. One, recommender systems research in the data mining literature has focused more on system design than understanding behavioral effects. Two, the marketing literature is just beginning to examine such behavioral effects. Three, of the existing behavioral work, the focus has been more on individual choice outcomes than aggregate effects. Four, regarding aggregate effects, there are opposing conjectures as to the impact of recommenders on sales diversity. This last point is the question we turn to now.

3. PROBLEM DEFINITION

This section sets the problem context, defines a measure of sales diversity, and formulates the question to be investigated. Our context is a market with a single firm selling one class of good (e.g., music versus movies). Within this one class, the firm can offer many items (e.g., CDs by thousands of artists).

3.1 Measure of Sales Diversity

Our outcome of interest is *sales diversity*. Before examining recommender systems’ effects, it is necessary to distinguish between sales and product diversity. *Product diversity*, or product variety, typically measure how many different products a firm offers. It is a supply-side measure of breadth. In contrast, we use *sales diversity* to reflect the concentration of consumer purchases conditional on firms’ assortment decisions. To measure sales diversity, we adopt the Gini coefficient. The Gini coefficient is a common measure of distributional inequality. It is nearly a hundred years old and has been applied to many problems (e.g., from income inequality in economics to classifier “lift” in data mining). For a more recent review, see Dorfman (1979).

Let $L(u)$ be the Lorenz curve denoting the percentage of the firm’s revenue generated by the lowest 100u% of goods sold during a fixed time period. The Gini coefficient is then defined as

$$G = 1 - 2 \int_0^1 L(u) du$$

Graphically this corresponds to $G = A/(A+B)$ in Figure 1. That is, the Gini coefficient is the ratio of A, the area between the 45° line and the Lorenz curve, to (A+B), the area below a Lorenz in which all products have equal sales (i.e. $L(u)=f(u)=u$). A similar formula can be applied for the discrete case as well. This gives $G \in [0,1]$. The value $G = 0$ reflects diversity (all products have equal sales), while values closer to 1 represent concentration (a small number of products account for the majority of sales).

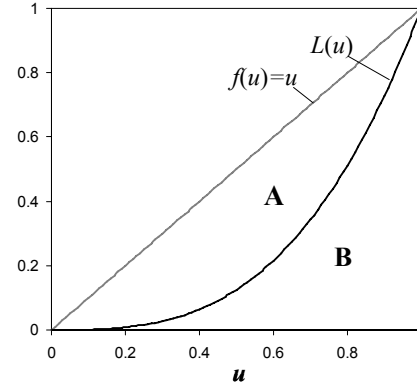


Figure 1. Lorenz curve

3.2 Problem Statement

Consider a firm with I customers c_1, \dots, c_I , J products p_1, \dots, p_J , and a sales database S . Define a recommender system as a function $r(c_i, S) \rightarrow p_j$ that maps a customer $c_i \in C$ and a database S onto a recommended product $p_j \in P$. Typically S is a matrix of users versus items with each element representing either number of purchases or a user rating. This paper focuses on recommenders that use purchase data. Both are used in industry, but the former appears more common because users are not always willing to provide ratings manually for catalog items.

Consider a set of different recommender systems r_1, \dots, r_k . Each r_i reflects certain design choices that distinguish it as a particular recommender system. For example, r_i might be a user-to-user collaborative filter, while r_j might be Amazon.com’s item-to-item collaborative filter. Denote by G_0 the Gini coefficient of the firm’s sales during a fixed time period in which a recommender system was *not* used. In contrast, let G_i be the Gini coefficient of the firm’s sales in which recommender system r_i was employed with all else equal. We assess the effect of a recommender system using the following definition.

Definition. Recommender bias. Recommender r_i is said to have a concentration bias, diversity bias, or no bias depending on the following conditions:

$$\begin{cases} \text{Concentration bias} & G_i > G_0 \\ \text{Diversity bias} & G_i < G_0 \\ \text{No bias} & G_i = G_0 \end{cases}$$

Correspondingly, the hypotheses we wish to test are

$$H_0: G_i = G_0 \text{ versus } H_1: G_i \neq G_0$$

where r_i is a specific, popular, real-world recommender. We examine these questions using an analytical model and simulation.

4. ANALYTICAL MODEL

This section presents an analytical model to explore a recommender system's effects. We first present a model representing a stylized collaborative filter. Afterward we derive properties of the model that reveal how the recommender affects behavior and market shares. These properties also allow us to calculate the net effect on sales diversity.

4.1 Assumptions

We consider a sequence of customers making purchases. Each customer's decision is which product to buy and does not include whether to buy. By assuming all consumers make purchases, we can isolate choice effects from incidence. As an example, this could reflect the choices of customers at a subscription media service. The firm's goal in such a case is to maximize customer satisfaction. Increasing satisfaction may either decrease churn or increase usage, which in turn can increase advertising revenue.

4.2 Preliminaries

Throughout this section, recommender system is synonymous with collaborative filter.¹ Collaborative filters can operate on purchase data or ratings data. Our model considers purchases. This choice corresponds to a prior hypothesis about the bias of such systems and is discussed later.

We now define how the recommender affects the choice process. The recommender's input is the firm's sales history. A history $H_t = \{(c,p)_1, \dots, (c,p)_t\}$ records which product $p \in \{p_1, \dots, p_J\}$ was purchased by which consumer $c \in \{c_1, \dots, c_t\}$ on each occasion up to t . Further, suppose a consumer c can be represented by a type vector $\theta_c = (\theta_{c,1}, \dots, \theta_{c,J})$ giving his purchase probabilities over the J items. As before, subscripts identify consumers c_1, \dots, c_t and products p_1, \dots, p_J . In addition, let superscripts identify the active customer and his type at t . For example, $c^t \in \{c_1, \dots, c_t\}$ is the active customer at t and θ^t is the type of c^t . The stylized recommender g and influence model f are defined by the following two-part process.

$$s^t = g(H_{t-1}, c^t) \tag{1}$$

$$H_t = f(s^t, \theta^t) \tag{2}$$

The recommender is a function g that maps a sales history H_{t-1} and active customer label c^t into a recommended product $s^t \in \{p_1, \dots, p_J\}$. The firm, of course, knows the customer label c^t for whom it is making a recommendation; the firm does not, however, know θ^t and thus can only use the information in H_{t-1} . Next, an influence function f describes the consumer's response: f stochastically maps the recommendation s^t and active customer's type θ^t into the choice outcome and, equivalently, the updated history H_t .

4.3 A Two-Urn Model for Recommenders

We now complete the model defined in (1) and (2) by specifying functional forms for g and f . For tractability, we make several

simplifying assumptions. First, we assume there are only two products w and b (white and black). Second, we pre-select a segment of consumers who are similar to one another and have the same type $\theta = \{p, 1-p\}$ as prior probabilities for w and b in the absence of recommendations. Third, we make g a function $g(X_t)$ instead of $g(H_{t-1}, c^t)$, where X_t is the market share of w just before purchase t . Regarding this third assumption, it is a common feature of many recommenders to ignore the time order in H_t , which is part of the reduction from H_{t-1} to X_t . It is a less common assumption to drop the customer labels in H_t ; we adopt this on the basis of having preselected a segment of similar consumers and so g treats them identically.

We describe the process as an urn model. Urn models are appealing because they have simple descriptions but can still model complex phenomena. Johnson & Kotz (1977) show how many significant results from probability theory can be derived from such simple models.

To specify (1) and (2), consider the two urn system of Figure 2. Urn 1 contains white and black balls representing products w and b . A fraction, p , of the balls in urn 1 are white. This fraction is the consumer's prior probability for w in the absence of recommendations. Urn 2 is the recommender: its contents reflect the sales history, and it produces recommendations according to a function $g(X_t)$, where X_t is the fraction of w in urn 2 just before t . As with many recommenders, for simplicity we ignore time order in H_t and use X_t instead. To start, urn 2 contains one w and one b . At time $t=1$, a ball is drawn with replacement from urn 1 representing the consumer's prior choice before seeing the recommendation. Next, a recommended product is drawn with replacement from urn 2 according to $g(X_t)$. With probability r , the consumer accepts this recommendation (the ball from urn 2), and with probability $(1-r)$ the consumer retains the original choice (the ball from urn 1). Thus, r is the strength of the recommender. The ball chosen represents the actual product purchased. Afterward, a copy of the ball chosen is added to urn 2, which is equivalent to updating the recommender's sales history. Consumer 2 then arrives, and the process repeats (p and r are the same, but X_2 is used instead of X_1). The process then repeats for the entire sequence of consumers.

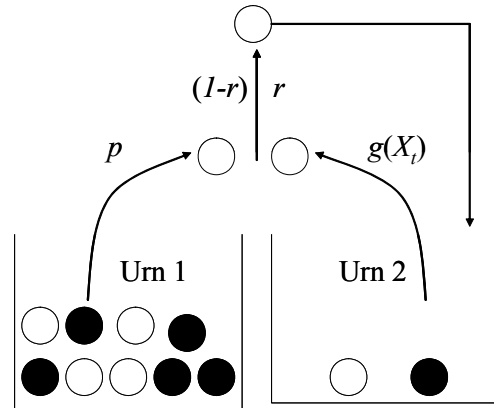


Figure 2. A two-urn model for recommender systems

For the recommender g , we use a sigmoid function and let the recommendations be given stochastically. Let

¹ Collaborative filters are commonly deployed in industry, but other types of recommenders exist (e.g., content-based). See Adomavicius & Tuzhilin (2005) for a taxonomy.

$$g(x) = 1/[1 + e^{-s(x-1/2)}] \quad (3)$$

for $x \in [0,1]$. This is an s-shaped curve centered at $x = 1/2$, and for large s it approximates a step function (Figure 3 shows an example with $s = 50$). Then the probability that urn 2 recommends w given the market share X_t is

$$P(\text{Urn 2 recommends } w \mid X_t = x) = g(x) \quad (4)$$

This choice of g means that the recommender generally suggests the majority share product. If more consumers have bought w , then w is generally recommended and vice-versa for b . This choice of g has a parallel with collaborative filters. Collaborative filters find similar user groups and recommend the most popular item within such groups; we assume, for simplicity, such a group has been preselected (consumers of type p) and that g operates solely in this group. Section 5 relaxes this assumption by allowing the recommender to operate across many types of consumers.

While collaborative filters are generally deterministic, our g is stochastic for two reasons. The primary reason is tractability. Whereas a step function might be an appropriate g (hard threshold), to derive model properties we need a continuous g and so use the sigmoid (soft threshold) instead. This is not too restrictive, since g can be made arbitrarily steep through its parameter s . Second, the noise introduced by stochastic g could be considered more realistic. If real recommenders have other pieces of user information that make them deviate on occasion from the majority-wins rule, this could be accommodated by the small amount of noise in g .

Returning to the consumer, after accounting for the prior preference p and recommender's influence we can define

$$\begin{aligned} f(X_t) &\equiv P\{w \text{ chosen on occasion } t \mid X_t\} \\ &= p(1-r) + g(X_t)r \\ &= p(1-r) + r/[1 + e^{-s(x-1/2)}] \end{aligned} \quad (5)$$

The function notation $f(X_t)$ serves two purposes. It relates the two-urn model to the two-part model of equations (1)-(2). Further, the function notation emphasizes the dependence on X_t , a dependence that we use next to derive results about the market outcomes.

4.4 Model Results

The market shares form a sequence X_1, X_2, \dots . Our interest is the limiting behavior of X_n for it reveals how a recommender affects shares that would otherwise be $\{p, 1-p\}$. A result from probability theory describes such behavior. The theorem is due to Hill et al. (1980); since this time, the results have been extended and applied elsewhere elsewhere in economics (Arthur, 1994).

Equation (5) specifies a continuous f that maps the unit interval into itself. This is so because $0 \leq p, g(x) \leq 1$ and $f(x)$ is a convex combination of the two. The input is the ingoing share X_t and the output $f(X_t)$ is the chance of w on occasion t . Let $C = \{x: f(x) = x\}$ and $C = U \cup D$ be the union of disjoint *upcrossings* and *downcrossings*. A point $x_0 \in [0,1]$ is an *upcrossing* (*downcrossing*) if, for all x in some neighborhood of x_0 , $x < x_0$ implies $f(x) < x$ ($f(x) > x$) and $x > x_0$ implies $f(x) > x$ ($f(x) < x$). By Hill et al.'s strong law for generalized urn processes (1980), $\{X_n\}$ converges almost surely to a random variable X with support D .

A graphical example helps to illustrate the results. For sake of illustration, take $p = 0.7$ and $r = 0.5$ for the consumer parameters and $s = 50$ for the sigmoid. Solving $f(x) = x$ gives

$$f(x) = x \leftrightarrow x \in C = \{0.35, 0.48, 0.85\}$$

As shown in Figure 4, the first and third points are downcrossings and the second is an upcrossing. Thus by the strong law, $X_n \rightarrow X$ where X has support $D = \{0.35, 0.85\}$. These outcomes are very different: indeed, the market leader reverses from one to the other.

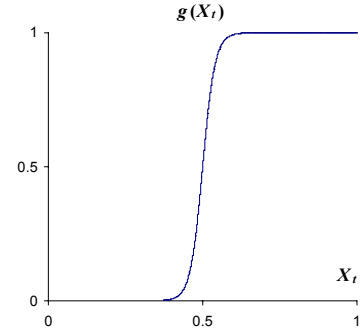


Figure 3. Recommender function $g(X_t)$

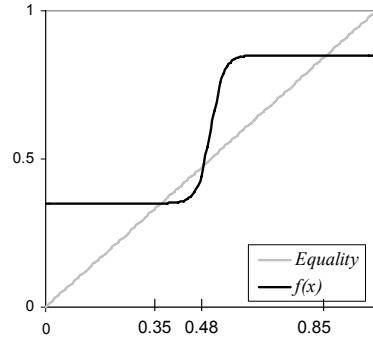


Figure 4. Urn function $f(x)$ for $p=0.7, r=0.5$

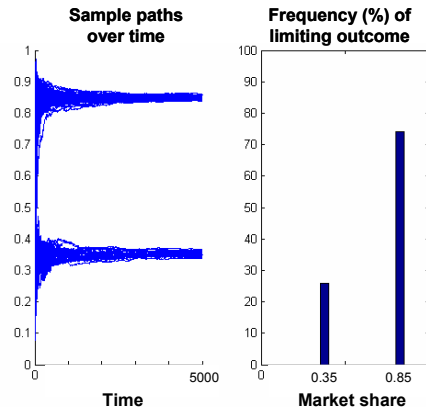


Figure 5. Two outcomes for our example $f(x)$

The strong law for urn processes gives the support of the limiting distribution, but it does not give the frequencies with which each outcome occurs. We use simulation to calculate these frequencies.

For the same example, we estimate that the shares $\{0.35, 0.85\}$ occur $\{26\%, 74\%\}$ of the time respectively (averages are over 1000 experiments of 5000 iterations each). This is illustrated graphically in Figure 5.

Returning to the main question of interest, we wish to know if recommendations have increased or decreased diversity. Let G_0 be the Gini without recommendations, G_f be the Gini under urn function f , and $G(a)$ be shorthand for the Gini of two products with shares $\{a, 1-a\}$. Then the difference in Gini coefficients before versus after recommendations is

$$\begin{aligned} E[G_f] - G_0 &= [0.26G(0.35) + 0.74G(0.85)] - G(0.7) \\ &= 0.26(0.15) + 0.74(0.35) - 0.20 \\ &= 0.30 - 0.20 \\ &> 0 \end{aligned}$$

Since $E[G_f] > G_0$, the recommender on average *increases* concentration. Although two outcomes are possible, one having higher concentration $G=0.35$ and one having lower concentration $G=0.15$, on average concentration is increased. The interpretation is as follows. Outcome 0.85 is more likely because preferences are strong ($p=0.7$) and the recommender is not too influential ($r=0.5$). Thus most outcomes (74%) are close to p . That the limiting share $0.85 > p$ reflects the recommender's ability to reinforce an already popular product's popularity. In a minority of cases (26%), outcome 0.35 is selected. Here, enough b occur early on that the recommender fixes on b .

This example corresponds to one particular choice of (p, r) . Other choices would lead to different urn functions $f(x)$, and thus different limiting support points (either 1 or 2 downcrossings at different positions in Figure 4), and thus different net effects on diversity (i.e. whether $E[G_f] - G_0 > 0$). For space reasons, we save for forthcoming work formal propositions relating (p, r) to the number of outcomes and quantity $E[G_f] - G_0$. Instead, as a concise summary, we present a visual result showing that for almost any (p, r) in our model, concentration on average increases (Figure 6). The exceptions are the minority and occur when consumers have strong preferences ($p \approx 0$ or $p \approx 1$) and the recommender is highly influential ($r \approx 1$). In such cases, the initial Gini is high (because of such strong preferences), but the recommender is so strong that it can push some outcomes away from the initial polarization.

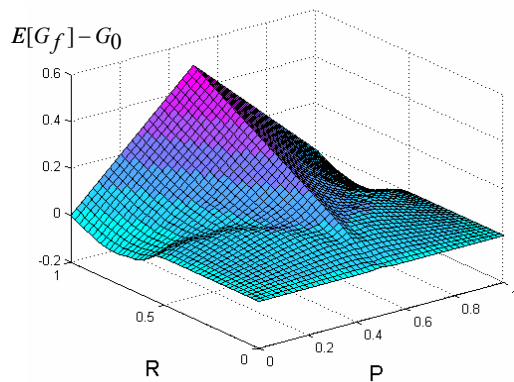


Figure 6. For almost all (p, r) , recommendations increase concentration as seen where $E[G_f] - G_0 > 0$.

To summarize the findings of this section, we find that in particular sample paths recommenders can increase or decrease

sales concentration. The example had one limiting outcome of each type. These limiting outcomes, however, do not necessarily occur equally often. In expectation, the Gini is likely to increase (more concentration) for almost all (p, r) . This has an important implication for empirical studies: studies of sales diversity at a single firm might be misleading. This is so because in one sample path either outcome could occur. To assess a recommender's net effect, our model suggests one must have data from an ensemble of firms to average over (and each using the same recommender). Such data requirements could be strong. The conclusions of this section are based on a stylized model of recommenders in a context with two products. To increase their realism and also test robustness, we now use simulation. to combine multi-product choice models with actual implementations of recommender systems.

5. SIMULATIONS

This section extends the analytical model. It uses simulation to combine a multi-product choice model with actual implementations of recommender systems.

5.1 Rationale for Simulation

Simulation offers three benefits for this problem. First, while actual recommenders are difficult to represent in an analytic model, they can be easily instantiated in simulation. Second, a challenge in analytic models is heterogeneity. Heterogeneity can enter in two ways: users can have different prior preferences over products and the recommender can respond differently to each user. Both are easily accommodated in simulation. Third, simulation solves the “ensemble” problem described above since one can replicate an experiment *in silico* as many times as needed.

5.2 Choice Model and Simulation Design

We investigate the sales diversity question by using an agent-based simulation that combines a consumer choice model with actual recommender systems. As before, we assume the number of products supplied is fixed, consumers are required to make purchases, and that repeat purchases are permitted. Examples of such contexts could include DVD rentals from Netflix or music streaming from a subscription service such as Rhapsody.

An overview of the process is as follows. Every period, each of I consumers c_1, \dots, c_I sequentially purchase one of J products p_1, \dots, p_J . To model choice among products, we use the multinomial logit. Our logit model compares all products to a consumer's ideal point in attribute space. Further, just before each consumer chooses a product, a recommendation p_k is generated using recommender r . The salience of the recommended product is increased by an amount δ in the logit model for that one occasion. The next consumer makes purchases in a similar manner, and after all consumers have purchased the process repeats. After a predetermined number of iterations, G_i is computed and compared to G_0 . Note, G_0 is pre-calculated in a burn-in period with the same process except that recommendations are not enabled. This amounts to a computer experiment in which ‘recommendations off’ is the control and ‘recommendations on’ is the treatment.

We now discuss each of the main simulation components: (i) the map of products and ideal points (ii) the recommender system r , (iii) the choice model, and (iv) the salience factor δ .

```

Input: maps of products and ideal points, salience factor  $\delta$ ,  $r$ 
For  $t = 1$  to TotalTime {
  For  $i = 1$  to  $I$  Customers {
    Firm uses  $r$  to generate recommendation  $p_k$  for customer  $i$ 
    Utility of  $p_k$  boosted by  $\delta$  temporarily
    Customer  $i$  uses a logit model to select a product
  }
}
Output  $G_i$ 

```

Figure 7. Simulation overview

(i) *Map of products and ideal points in attribute space.* The map of products and ideal points is the input for each consumer's choice model. Plotting ideal points and product locations goes back at least to Hotelling (1929). While Hotelling's line segment market was analytical, work in Marketing has estimated maps empirically to understand market structure (e.g., Elrod, 1988). Our consumers and products are each points in a two-dimensional attribute space. The use of two dimensions is for simplicity and visualization. For contexts in which more than two attributes are relevant, our maps can be considered dimensionality-reduced versions, as is common in Marketing research. We take both ideal points and products to have bivariate normal distributions with identity covariance. The normality assumption for consumers has been used often in factor-analytic market maps (e.g., Elrod 1988; Elrod & Keane, 1996). In an extended, forthcoming version of this work, we test sensitivity to other distributions, but it does not seem to affect the conclusions drawn. A sample map with 50 ideal points and 50 products appears in Figure 8.

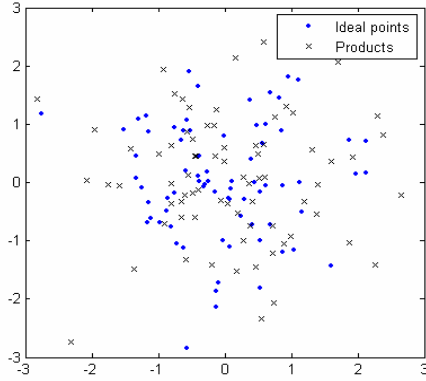


Figure 8. Map of products and ideal points in attribute space

(ii) *The recommender system.* A primary advantage of simulation is the ability to test real recommender systems. We study sales diversity under two recommender systems, termed here r_1 and r_2 . In the taxonomy of Adomavicius and Tuzhilin (2005), both are memory-based, collaborative filters based on cosine similarity.

Recommender r_1 is the most basic collaborative filter: for a given user, it finds the set N of the n most similar customers by using cosine similarity to compare two vectors of purchase counts. It then recommends the most popular item among this group. For a given user, the system recommends product

$$r_1: j^* = \arg \max_j \sum_{i \in N} sales_{ij} \quad (6)$$

where $sales_{ij}$ denotes the number of times customer c_i purchased product j . Recommender r_2 is the same except for one important difference. When selecting the most popular product among the similar users, candidate items are first discounted by their overall popularity. Let f_j denote product j 's sales among all consumers. Then for a given user, r_2 would recommend

$$r_2: j^* = \arg \max_j \frac{1}{f_j} \sum_{i \in N} sales_{ij} \quad (7)$$

The motivation for r_2 's popularity discounting is a belief that popular items are so obvious that they should not be suggested. For example, if a consumer is expected to buy or be aware of a product with high probability, the firm should recommend something *else* if it wishes to generate incremental sales. Interviews with industry experts suggested such popularity discounting is common practice. Note, r_2 is not the same as applying *tf-idf* weights to algorithm r_1 . Using *tf-idf* would insert discounting in the user similarity calculation (Breese et al., 1998), whereas r_2 computes an undiscounted user similarity and discounts popularity in the final argmax of equation (7). (In a longer, forthcoming version of this work, we show results for *tf-idf* as well. The results are directionally the same.)

(iii) *Choice model.* At each step of the simulation, a consumer purchases one of the items in his choice set. We model choice using the multinomial logit. The logit is well established in economics and marketing and has an axiomatic origin in random utility theory (for an application, see Guadagni & Little, 1983). Consumer i 's utility for product j defined is $u_{ij} = v_{ij} + \epsilon_{ij}$, where v_{ij} is a deterministic component and ϵ_{ij} is an i.i.d. random variable with extreme value distribution. Under this, it has been shown that

$$P(\text{Consumer } i \text{ buys } j) = \frac{e^{v_{ij}}}{\sum_{k=1}^J e^{v_{ik}}} \quad (8)$$

where the summation is over all J items. The deterministic utility v_{ik} is often modeled as a linear combination of a brand intercept, product attributes, and market-related covariates (e.g., price, promotion). In our context, since all relevant variables are encompassed in the map of products and ideal points, we define the deterministic utility of product p_j to consumer c_i as

$$v_{ij} \equiv \text{similarity}_{ij} = -k \log \text{distance}_{ij} \quad (9)$$

where distance_{ij} is Euclidean distance between consumer i and product j on the map. Since the map measures distances, some such monotonic transformation is needed to relate distance and similarity. The log transformation chosen is commonly applied in marketing (e.g., Schweidel et al., 2006). For forthcoming work, we have tried other transformations, and they do not appear to affect the results. The parameter k determines the breadth of consumer preference: high k means the consumer prefers only the closest products, whereas low k means the consumer weighs far and near products more evenly (i.e. distance is more important for higher k and vice-versa). To chose k , we obtained data from a large CD retailer for which the Gini was roughly 0.6. Using this, we increased k from 0 and stopped when the implied Gini of our map was close to 0.6. This occurred at $k=10$. The experiments

below use this value. A forthcoming version of this work shows sensitivity to k , but the results do not appear to change direction.

(iv) *Saliency* δ . The term δ is the amount by which a recommended product’s saliency is temporarily increased in the consumer’s choice set. If product j is recommended to i , the deterministic utility becomes $v_{ij} \leftarrow v_{ij} + \delta = -k \log \text{distance}_{ij} + \delta$. This increased saliency lasts for the current choice occasion only (unless p_j happens to be recommended in a future iteration, at which point its saliency will again be temporarily boosted). We initially chose $\delta = 15$. For a typical consumer i on our maps with $k=10$, i ’s 75th percentile similarity item has nearly zero probability versus the item closest to i . Setting $\delta = 15$ instantiates a rule of thumb that when i ’s 75th percentile item is recommended, it becomes about even in probability with the most similar item. Again, in longer, forthcoming work we show sensitivity to δ . The strength of results diminishes as expected as $\delta \rightarrow 0$ since the recommender is less influential in such cases.

6. RESULTS

We now present simulation results for the two real-world recommenders. We use 50 ideal points and 50 products sampled bivariate normal $N_2(\mathbf{0}, \mathbf{I})$, $k = 10$ and $\delta = 15$ as discussed above. Each simulation is 100 iterations with recommendations off and 100 iterations with recommendations on.

6.1 Example of a Single Sample Path

Before presenting overall results, we show one sample path as an example. At first, recommendations are disabled and customers make purchases for 100 periods. Then r_1 is enabled, customers make purchases for another 100 periods, and the Ginis before and after are compared. The Lorenz curves from both periods and corresponding Ginis are shown in Figure 9.

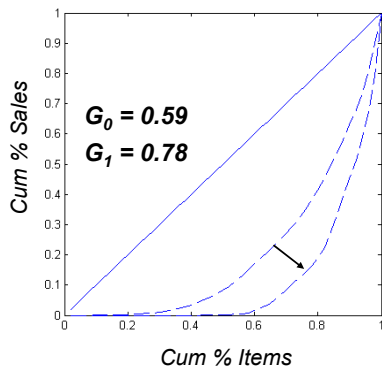


Figure 9. An example run: before versus after for r_1

The example shows $G_1 - G_0 = 0.78 - 0.59 = 0.19 > 0$, and hence r_1 increases concentration in this run. This is for a single sample path, and below we conduct a more systematic comparison.

6.2 Simulation Results

We now analyze the results for a large number of simulation runs. Using the same parameters as above, we use 1000 experiments/maps each for r_1 and r_2 . Results appear in Table 1 with standard errors given in parentheses. The experiments show the following. First, both recommenders on average have a homogeneity bias, as reflected by $G_1 > G_2 > G_0$. The “standard” collaborative filter r_1 shows the greatest bias. The popularity-discounting recommender

r_2 shows evidence of concentration bias too. It is not surprising that $G_1 > G_2$. However, we do find it surprising that $G_2 > G_0$: a priori we could not rule out the possibility of r_2 ’s popularity discounting leading to a reduction in diversity. In fact, in a small number of runs (17%) r_2 increases diversity, but in the majority of runs (83%) and on average it reduces diversity. A test of paired differences (pre vs. post recommendations) for the 1000 experiments shows all results to be significant at the 0.05 level.

Table 1. Comparison of r_1 and r_2 (1000 experiments)

	r_1	r_2
Gini before: G_0	0.533 (.045)	0.533 (.046)
Gini after: G_i	0.782 (.031)	0.567 (.047)
Difference	+0.249	+0.034
p -value	$< 10^{-6}$	$< 10^{-6}$

6.3 Discussion of Results

We attribute the bias of r_1 primarily to its use of implicit voting, which takes purchases as implied positive ratings and non-purchases as implied negative ratings (Breese et al., 1998). This creates a self-reinforcing cycle of popularity: frequently purchased items are recommended more and items recommended more are more likely to be purchased. The concentration bias of r_2 we also attribute to implicit voting. With r_2 , however, the bias is moderated by popularity discounting. While we expected $G_2 < G_1$, we were surprised to see on average $G_2 > G_0$. Even though r_2 uses implicit votes, it explicitly discounts popular items, so beforehand we could not rule out the possibility of its decreasing concentration. Ultimately, this was not the case for its average effect. For both recommenders, we have conducted a sensitivity analysis to appear in a longer version of this work. In that work we also test other recommender systems as well as modifications to the consumer utility specification.

7. CONCLUSIONS

Despite the prevalence of recommender systems, there is still much we do not know about how they affect behavior. This paper examined their effect on buying behavior and offered initial evidence that recommender systems do influence sales diversity. A simple collaborative filter was investigated analytically. We found that for most model settings, average concentration increased, although for specific sample paths concentration could increase or decrease. Using simulation, two popular collaborative-filtering based systems were investigated: one was expected to have a concentration bias and the other a diversity bias, but both, surprisingly, were found to exert a concentration bias on average. It remains to be seen which, if any, real-world design choices lead to a diversity bias in expectation. Consistent with the analytical model, our simulations also demonstrate that the realized sales diversity is path dependent. Within particular sample paths, the same recommender can create diversity in one case and concentration in another. Finally, the simulations demonstrate how basic design choices affect the outcome and that some designs may be associated with greater bias than others.

Future Work: For the analytical model, we are exploring ways to incorporate heterogeneity in the consumer's urn and recommender function g . We are also exploring the multi-product case. For the simulation, we are pursuing several new directions.

1) *Utility specification – Variety seeking and loyalty.* In forthcoming work, we extend the consumer's utility to include variety and loyalty. Since our context allows repeat purchases, this aspect controls the degree to which consumers prefer choices to be different (variety seeking) versus similar (loyalty) over time.

2) *Utility specification – Awareness.* In forthcoming work, we allow that consumers are no longer assumed initially aware of all products, and recommendations serve as a device by which they gradually learn and become aware of the other items (Bodapati).

3) *Additional recommender systems.* In forthcoming work we expand the set of recommender systems tested.

4) *Welfare implications and externalities.* Thus far we have asked how recommender systems affect diversity: do they push the Lorenz curve inward versus outward? We have not asked about the welfare implications of such effects: does such a movement of the curve leave consumers better versus worse off? The latter is beyond the scope of this work but is an important question that has implications for both new system design and understanding the consequences of (tacit) design choices in existing systems.

A further interesting aspect arises to the extent that externalities exist for media goods. If, for example, there is a benefit to reading popular books or seeing popular movies (e.g., because it increases the likelihood of being able to discuss the experience with others), then consumer utility involves a tradeoff between a Hotelling-like similarity and the externality from consuming a popular product. To the extent that such externalities are strong, it would be interesting to see if they pose a limit, or upper bound, on the degree of diversity consumers would ever prefer. We hope to explore these questions in future work as well.

8. ACKNOWLEDGMENTS

The authors thank Eric Bradlow, Terry Elrod, Pete Fader, David Schweidel, Michael Steele, three anonymous reviewers, and OPIM and WISE seminar participants for their helpful comments.

9. REFERENCES

- [1] Adomavicius, G. and A. Tuzhilin. 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *TKDE* 17(6):734-749.
- [2] Anderson, C. 2004. The long tail. *Wired Magazine* 12:10.
- [3] Arthur, W. B. 1994. *Increasing Returns and Path Dependence in the Economy*. Ann Arbor: U. of Michigan.
- [4] Bikhchandani, S., D. Hirshleifer, and I. Welch. 1998. Learning from the behavior of others: conformity, fads, and informational cascades. *J. Econ. Perspectives* 12:3 151-170.
- [5] Bodapati, A. Recommendation systems with purchase data. *Journal of Marketing Research*, forthcoming.
- [6] Breese, J., D. Heckerman, and C. Kadie. 1998. Empirical analysis of predictive algorithms for collaborative filtering. In *Proc. of the 14th Conference of Uncertainty in AI*.
- [7] Brynjolfsson, E., Y. Hu, and M. Smith. 2006. From niches to riches: the anatomy of the long tail. *Sloan Management Review* 47(4) 67-71.
- [8] Clemons, E.K., G.G. Gao, and L.M. Hitt. 2006. When online reviews meet hyperdifferentiation. *J. of MIS* 23(2):149-171.
- [9] Cooke, A., H. Sujan, M. Sujan, and B. Weitz. 2002. Marketing the unfamiliar: the role of context and item-specific information in electronic agent recommendations. *Journal of Marketing Research* 39(4):488-497.
- [10] Dorfman, R. 1979. A formula for the Gini coefficient. *The Review of Economics and Statistics* 61(1):146-149.
- [11] Elrod, T. 1988. Inferring an ideal-point product-market map from consumer panel data. In *Data, Knowledge and Decisions*, p. 240-249. Edited by W. Gaul and M. Schader. Berlin: Springer-Verlag.
- [12] Elrod, T. and M. P. Keane. 1995. A factor-analytic probit model for representing the market structure in panel data. *J. of Marketing Research* 32(1):1-16.
- [13] Guadagni, P.M. and J.D.C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Science* 2(3):203-238.
- [14] Hill, B. M., D. Lane, and W. Sudderth. 1980. A strong law for some generalized urn processes. *Annals of Probability* 8(2):214-226.
- [15] Hotelling, H. 1929. Stability in competition. *The Economic Journal* 39:41-57.
- [16] Johnson, N. L. and S. Kotz. 1977. *Urn Models and their Applications*. New York: John Wiley and Sons.
- [17] Linden, G., B. Smith, and J. York. 2003. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing* 7(1):76-80.
- [18] Mooney, R. J. and L. Roy. 2000. Content-based book recommending using learning for text categorization. *5th ACM Conf. on Digital Libraries*, p.195-204.
- [19] Pennock, D., E. Horvitz, and C. Lee Giles. 2000. Social choice theory and recommender systems: analysis of the axiomatic foundations of collaborative filtering. *17th NCAI*.
- [20] Pham, A. and J. Healey. 2005. Tell you what you like. *Los Angeles Times*, 20 September.
- [21] Sarwar, B., G. Karypis, J. Konstan, and J. Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *Proc. of the 10th International World Wide Web Conference*.
- [22] Schafer, J., J. Konstan, and J. Riedl. 1999. Recommender systems in e-commerce. *1st ACM EC*, p.158-166.
- [23] Schein A. I., A. Popescul, L. H. Ungar, and D. M. Pennock. 2002. Methods and metrics for cold-start recommendations. *SIGIR*, p. 253-260.
- [24] Schweidel, D., E.T. Bradlow, and P.S. Fader. 2006. Modeling the evolution of customers' service portfolios from a multi-service provider. Working paper.
- [25] Senecal, S. and J. Nantel. 2004. The influence of online product recommendations on consumers' online choices. *Journal of Retailing* 80:159-169.