Towards a Theory Model for Product Search*

Beibei Li
bli@stern.nyu.edu

Anindya Ghose
aghose@stern.nyu.edu

Panagiotis G. Ipeirotis
panos@stern.nyu.edu

Department of Information, Operations, and Management Sciences
Leonard N. Stern School of Business, New York University
New York, New York 10012, USA

ABSTRACT

With the growing pervasiveness of the Internet, online search for products and services is constantly increasing. Most product search engines are based on adaptations of theoretical models devised for information retrieval. However, the decision mechanism that underlies the process of *buying a product* is different than the process of *locating relevant documents or objects*.

We propose a theory model for product search based on expected utility theory from economics. Specifically, we propose a ranking technique in which we rank highest the products that generate the highest *surplus*, after the purchase. In a sense, the top ranked products are the "*best value for money*" for a specific user. Our approach builds on research on "demand estimation" from economics and presents a solid theoretical foundation on which further research can build on. We build algorithms that take into account consumer demographics, heterogeneity of consumer preferences, and also account for the varying price of the products. We show how to achieve this *without knowing* the demographics or purchasing histories of *individual* consumers but by using aggregate demand data. We evaluate our work, by applying the techniques on hotel search. Our extensive user studies, using more than 15,000 user-provided ranking comparisons, demonstrate an overwhelming preference for the rankings generated by our techniques, compared to a large number of existing strong state-of-the-art baselines.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Algorithms, Economics, Experimentation, Measurement

Keywords

Consumer Surplus, Economics, Product Search, Ranking, Text Mining, User-Generated Content, Utility Theory

1. INTRODUCTION

Online search for products is increasing in popularity, as more and more users search and purchase products from the Internet. Most search engines for products today are based on *models of relevance* from "classic" information retrieval theory [22] or use variants of faceted search [27] to facilitate browsing. However, the decision mechanism that underlies the process of *buying a product* is different from the process of finding a *relevant* document or object. Customers do not simply seek to find something relevant to their search, but also try to identify the "best" deal that satisfies their specific desired criteria. Of course, it is difficult to quantify the notion of "best" product without trying to understand what the users are optimizing.

Today's product search engines provide only rudimentary ranking facilities for search results, typically using a single ranking criterion such as name, price, best selling (volume of sales), or more recently, using customer review ratings. This approach has quite a few shortcomings. First, it ignores the *multidimensional* preferences of consumers. Second, it fails to leverage the information generated by the online communities, going beyond simple numerical ratings. Third, it hardly accounts for the *heterogeneity* of consumers. These drawbacks highly necessitate a recommendation strategy for products that can better model consumers' underlying behavior, to capture their multidimensional preferences and heterogeneous tastes.

Recommender systems [1] could fix some of these problems but, to the best of our knowledge, existing techniques still have limitations. First, most recommendation mechanisms require consumers' to log into the system. However, in reality many consumers browse only anonymously. Due to the lack of any meaningful, personalized recommendations, consumers do not feel compelled to login before purchasing. For example, on Travelocity, less than 2% of the users login. But even when they login, before or after a purchase, consumers are reluctant to give their individual demographic information due to a variety of reasons (e.g., time constraints or privacy issues). Therefore, most context information is missing at the individual consumer level. Second, for goods with a *low purchase frequency* for an individual consumer, such as hotels, cars, real estate, or even electronics, there are few repeated purchases we could leverage towards building a predictive model (i.e., models based on collaborative filtering). Third, and potentially more importantly, as privacy issues become increasingly important, marketers may not be able to observe the individual-level purchase history of each consumer (or consumer segment). In contrast, aggregate purchase statistics (e.g., market share) are easier to obtain. As a consequence, many algorithms that rely on knowing individual-level behavior lack the ability of deriving consumer preferences from such aggregate data.

Alternative techniques try to identify the "Pareto optimal" set of results [3]. Unfortunately, the feasibility of this approach diminishes as the number of product characteristics increases. With more than five or six characteristics, the probability of a point being classified as "Pareto optimal" dramatically increases. As a consequence, the set of Pareto optimal results soon includes *every* product.

So, how to generate the "best" *ranking* of products when

*This work was supported by NSF grants IIS-0643847 and IIS-0643846.

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.
WWW 2011, March 28–April 1, 2011, Hyderabad, India.
ACM 978-1-4503-0632-4/11/03.

consumers use multiple criteria? For this purpose, we use two fundamental concepts from economics: *utility* and *surplus*. Utility is defined as a measure of the relative satisfaction from, or desirability of, consumption of various goods and services [17]. Each product provides consumers with an overall utility, which can be represented as the aggregation of weighted utilities of individual product characteristics. At the same time, the action of purchasing deprives the customer from the utility of the money that is spent for buying the product. With the assumption consumers rationality, the decision-making process behind purchasing can be viewed as a process of utility maximization that takes into consideration both product quality and price.

Based on utility theory, we propose to design a new ranking system that uses demand-estimation approaches from economics to generate the *weights* that consumers implicitly assign to each individual product characteristic. An important characteristic of this approach is that *it does not require purchasing information for individual customers* but rather relies on aggregate demand data. Based on the estimated weights, we then derive the *surplus* for each product, which represents how much extra utility one can obtain by purchasing a product. Finally, we rank all the products according to their surplus. We extend our ranking strategy to a personalized level, based on the distribution of consumers' demographics.

We instantiate our research by building a search engine for hotels, based on a unique data set containing transactions from Nov. 2008 to Jan. 2009 for US hotels from a major travel web site. Our extensive user studies, using more than 15000 user evaluations, demonstrate an overwhelming preference for the ranking generated by our techniques, compared to a large number of existing strong baselines.

The major contributions of our research are the following:

- We aim at making recommendations based on better understanding of the underlying "*causality*" of consumers' purchase decisions. We present a user model that captures the decision-making process of consumers, leading to a better understanding of consumer preferences. This is in contrast to building a "black-box" style predictive model using machine learning algorithms. The causal model relaxes the assumption of a "consistent environment" across training and testing data sets and allows for changes in the modeling environment and predicts what *should* happen even when things change.
- We infer *personal* preferences from *aggregate* data, in a privacy-preserving manner. Our algorithm learns consumer preferences based on the largely anonymous, publicly observed *distributions of consumer demographics* as well as the observed aggregate-level purchases (i.e., anonymous purchases and market shares in NYC and LA), not by learning from the identified behavior or demographics of each individual.
- We propose a ranking method using the notion of *surplus*, which is not only theory-driven but also generates systematically better results than current approaches.
- We present an extensive experimental study: using six hotel markets, 15000 user evaluations, and using *blind tests*, we demonstrate that the generated rankings are significantly better than existing approaches.

The rest of the paper is organized as follows. Section 2 gives the background. Section 3 explains how we estimate the model parameters, specifically how we compute the weights associated with product characteristics. Section 4 discusses how we build our basic rankings, and how we can personalize the presented results. Section 5 provides the setting for the experimental evaluation, and Section 6 discusses the results. Finally, Section 7 discusses related work and Section 8 concludes.

2. THEORY MODEL

In this section, we provide the background economic theory that explains the basic concepts behind our model. We start by formalizing our problem and introducing the "economic view" of consumer rational choices. For better understanding, we introduce the following theoretical bases: *utility* theory, *characteristics-based* theory, and *surplus*.

2.1 Problem Description

In general, our main goal is to *identify the best products for a consumer*. The example illustrates this:

EXAMPLE 1. *Alice is looking for a hotel in New York City. She prefers a place of good quality but* preferably *costing not more than \$300 per night. She conducts a faceted search (e.g., with respect to* price *and* ratings*): Unfortunately, with explicit price constrain, she may miss some "great deal" with much higher value but a slightly higher price. For instance, the 5-star Mandarin hotel happens to run a promotion that week with a discounted price of \$333 per night. With the most luxurious environment and room services, the price for Mandarin would normally be around \$900 per night otherwise. So, although the price is \$33 above her budget, Alice would certainly be willing to "grab the deal" if this hotel appeared in the search result.*

However, the problem is how can Alice know that such a deal exists? In other words, how can we improve the search so that it can help Alice identify the "best value" products? To examine this problem, we introduce the concept of *surplus* from economics. It is a measure of the benefits consumers derive from the exchange of goods [17]. If we can derive the surplus from each product, then by ranking the products according to their surplus, we can easily find the best product that provides the highest benefits to a consumer. Now the question is, how to derive the surplus so that we can *quantify* the gain from buying a product? To do so, we introduce another concept: *utility*.

2.2 Choice Decisions and Utility Maximization

Surplus can be derived from *utility* and *rational choice* theories. A fundamental notion in utility theory is that each consumer is endowed with an associated utility function U, which is "*a measure of the satisfaction from consumption of various goods and services*" [17]. The rationality assumption defines that each person tries to maximize its own utility.[1]

In the context of purchasing decisions, we assume that the consumer has access to a set of products, each product having a price. Informally, buying a product involves the exchange of money for a product. Therefore, to analyze the purchasing behavior we need two components for the utility function:

- *Utility of Product:* The utility that the consumer will *gain* by buying the product, and
- *Utility of Money:* The utility that the consumer will *lose* by paying the price for that product.

In general, a consumer buys the product that maximizes utility, and does so only if the utility gained by purchasing the product is higher than the corresponding, lost, utility of the money.

More formally, assume that the consumer has a choice across n products, and each product X_j has a price p_j.[2] Before the purchase, the consumer has some disposable income I that

[1] While in reality consumers are not always rational, it is a convenient modeling framework that we adopt in this paper. As we demonstrate in the experimental evaluation, even imperfect theories generate good experimental results.

[2] To allow for the possibility of not buying anything, we also add a dummy product X_0 with price $p_0 = 0$, which corresponds to the choice of not buying anything.

generates a money utility $U_m(I)$. The decision to purchase X_j generates a product utility $U_p(X_j)$ and, simultaneously, paying the price p_j decreases the money utility to $U_m(I - p_j)$. Assuming that the consumer strives to optimize its own utility, the purchased product X_j is the one that gives the highest increase in utility.

This approach naturally generates a ranking order for the products: The products that generate the highest increase in utility should be ranked on top. Thus, to compute the increase in utility, we need the gained *utility of product* $U_p(X_j)$ and the lost *utility of money* $U_m(I) - U_m(I - p_j)$.

2.2.1 Utility of Product

Modeling the utility of a product can be traced back to Lancaster's *characteristics theory* [15] and Rosen's *hedonic price model* [24]. The hedonic price model assumes that differentiated products are described by vectors of objectively measured characteristics. In addition, the utility that a consumer has for a product can be decomposed into a set of utilities for each product characteristic. According to this model, a product X with K features can be represented by a K-dimensional vector $X = \langle x^1, \ldots, x^K \rangle$, where x^k represents the amount or quality of the k-th characteristic of the product. The overall utility of product X is then modeled by the function $U_p(x^1, ..., x^K)$.

One of the critical issues in this model is how to estimate the aggregated utility from the individual product characteristics. Based on the hedonic price model, we assume that each product characteristic is associated with a *weight* representing consumers' desirability towards that characteristic. Under this assumption, we further refine the definition of overall utility to be the aggregation of *weighted utilities* from the *observed* individual characteristics and an *unobserved* characteristic ξ:

$$U_p(X) = U_p(x^1, \ldots, x^K) = \sum_{k=1}^{K} \beta^k \cdot x^k + \xi, \qquad (1)$$

where β^k represents the corresponding weight that the consumer assign to the k-th characteristic x^k. Notice that with ξ we capture the influence of all product characteristics that are not explicitly accounted in our model. So, a product that consumers perceive as high-quality due to a characteristic not explicitly captured in our measurements (e.g. brand name), will end up having a high value of ξ.

2.2.2 Utility of Money

Given the utility of a product, to analyze consumers' motivation to trade money for the product, it is also necessary to understand the utility of money. This concept is defined as consumers' *happiness* for owning monetary capital. Based on Alfred Marshall's well-established principles [17], utility of money has two basic properties: *increasing* and *concave*.

- *Increasing*: An increase in the amount of money will cause an increase in the utility of money. In other words, the more money someone has, the better.
- *Concave*: The increase in utility, or *marginal utility* of money, is diminishing as the amount of money increases.

Based on these properties, an example of the utility function for money is shown in Figure 1. Note that with the concave form of the utility function, the slope is decreasing hence the marginal utility of money is diminishing. So, \$100 is more important for someone with \$1000 than for someone with \$100,000. This also implies that consumers are risk-averse under normal circumstances. This is because with the same probability to win or lose, losing N dollars in the assets will cause a drop in the utility larger than the boost while winning N dollars.

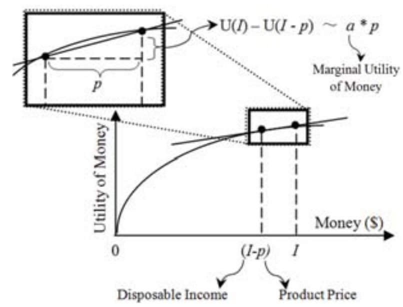

Figure 1: A concave, bounded, increasing function for "1utility of money," approximated with a linear function for small changes

The concave assumption can be relaxed when the changes in money are small. For most transactions, we often assume that the marginal utility of money is approximately *constant*. More formally, we assume that a consumer with income I receives a money utility $U_m(I)$. Paying the price p decreases the money utility to $U_m(I - p)$. Assuming that p is relatively small compared to the disposable income I, the marginal utility of money remains mostly constant in the interval $[I - p, I]$ [17]. Under this assumption, the utility of money that the consumer will lose by paying the price p for product X, can be thereby represented in a quasi-linear form as follows:

$$U_m(I) - U_m(I - p) = \alpha \cdot I - \alpha \cdot (I - p) = \alpha(I) \cdot p, \qquad (2)$$

where $\alpha(I)$ denotes the marginal utility for money for someone with disposable income I.

2.2.3 Challenges

Given the utility of product and utility of money, we can now derive the *utility surplus* as the increase in utility, or excess utility, after the purchase. More formally, the mathematical definition for utility surplus is provided as follows.

Definition 1.**:** The *utility surplus* (US), for a consumer with disposable income I, when buying a product X priced at p, is the *gain* in the utility of product U_p minus the *loss* in the utility of money U_m.

$$\begin{aligned} US &= U_p(X) - [U_m(I) - U_m(I - p)] + \varepsilon_j^i \\ &= \underbrace{\sum_k \beta^k \cdot x^k + \xi}_{\text{Utility of product}} - \underbrace{\alpha \cdot p}_{\text{Utility of money}} + \underbrace{\varepsilon}_{\text{Stochastic error}} \qquad (3) \end{aligned}$$

Note that ξ is a *product-specific* disturbance scalar summarizing unobserved characteristics of product X, and ε is a stochastic error term that is assumed to be i.i.d. across *products and consumers* in the selection process and is usually assumed to follow a Type I extreme-value distribution. □

In this theory model, the key challenge is to estimate the corresponding weights assigned by consumers towards money and product dimensions. We discuss this next.

3. ESTIMATION OF MODEL PARAMETERS

In the previous section, we have introduced the background of utility theory, characteristics-based theory, and surplus. Recall that our main goal is to identify the best product (with the highest surplus) for a consumer. This is complicated by the fact that utility, and therefore surplus, of consumers is *private* *and not directly observable.* As a result, there exists no "true observed utility" that we can compare with our "model predicted utility." Instead we need to observe the *behavior* of consumers and estimate the values of these latent parameters that explain best the consumer behavior. Furthermore, since we cannot assume that we can observe in detail the behavior of *individual* consumers, nor can we explicitly ask each consumer for their personal "tastes" (e.g., choice of a product, "weight" assigned to a product feature, etc.), we need to extract utilities and derive individual preferences by using *aggregate* data.

The basic idea is the following: If we know the utilities of different products for a consumer, we can estimate the demand for different products, as consumers will behave according to their utility-encoded preferences. So, if we observe the *demand* for various products, we can infer the preferences of the consumer population for different product aspects. Observing product demand is easier than it sounds: For example, we can observe *salesrank* on Amazon.com and transform salesrank to demand [8], or we can *directly observe the transactions* at marketplaces such as eBay and Amazon [11], or we can simply get directly anonymous transactions from a merchant.

In this section, we discuss how we estimate the parameters using aggregate demand data. First, in Section 3.1 we discuss the simpler case where consumers are homogeneous and have similar preferences. Then in Section 3.2 we analyze the more realistic case where consumers have different preferences, which depend on their demographics and purchase context.

3.1 Homogeneous Consumers: Logit Model

3.1.1 Model Specification

The basic Logit model, introduced by McFadden [18, 19], assumes that consumers have "homogeneous preferences" towards product characteristics. In other words, *the weights β and α are common across all consumers.* Thus, following Definition 1, the utility surplus for consumer i and product j is written as:

$$US_j^i = V_j(\alpha, \beta) + \varepsilon_j^i. \tag{4}$$

where $V_j(\alpha,\beta) = \sum_k \beta^k \cdot x_j^k + \xi_j - \alpha \cdot p_j$. Notice that we separate preferences towards product j, captured by $V_j(\alpha,\beta)$, from non-deterministic aspects of individual consumer behavior, captured by the error term ε_j^i.

According to the assumption of consumer rationality for utility maximization, the consumer chooses the product that maximizes utility surplus. Note that the choice is stochastic, given the error term ε_j^i. Therefore, in our scenario, the probability that a consumer i chooses product j is:

$$\begin{aligned} P(choice_j^i) &= P(US_j^i > US_l^i) \\ &\quad (\forall l \textit{ in the same market}, \ l \neq j)\,. \end{aligned} \tag{5}$$

Solving this equation, we have [18, 19]:

$$P(choice_j^i) = \frac{\exp\left(V_j(\alpha,\beta)\right)}{1 + \sum_l \exp\left(V_l(\alpha,\beta)\right)}. \tag{6}$$

In the homogeneous case, all consumers have the same α and β and this probability is *proportional* to the market share[3] of project j (the consumer-specific error term ε_j^i has disappeared).

Notice that the problem of estimating preferences can be now expressed as a logistic regression problem. What is worthwhile to mention is that this solution is not an adhoc choice, but is a direct derivation from a theory-driven user behavior model. Daniel McFadden got the Nobel prize in Economics in 2000 for establishing the connection between logistic regression and models of discrete user choice.

3.1.2 Estimation Methodology

Given Equation 6, we can estimate consumer preferences (expressed by the parameters α and β), by observing market shares of the different products. One challenge is that we need to know the "demand" for the "buy nothing" option in order to estimate properly the value $P(choice_j)$ in Equation 6.

Specifically, we set $P(choice_j) = d_j^{obs}/d_{total}$, where d_j^{obs} is the observed demand for product j and d_{total} is "total demand," which includes the demand for the buy-nothing option.[4] Taking logs in Equation 6 and solving the system [5]:

$$\ln(d_j^{obs}) = -\alpha \cdot p_j + \sum_k \beta^k \cdot x_j^k + \xi_j. \tag{7}$$

Such a model can be easily solved to acquire the parameters β and α using any linear regression method, such as ordinary least squares (OLS).

EXAMPLE 2. *Assume that we have a hotel market in New York, with two hotels: Hotel M (Mandarin Oriental, 5-star), and Hotel D (Doubletree, 3-star). From day 1 to 3, we observe that the price for Mandarin Oriental is \$500, \$480 and \$530 per night. We also observe a corresponding demand of 400, 470, and 320 bookings, respectively. Meanwhile, the price for Doubletree is \$250, \$270 and \$225 per night, and its corresponding demand is 600, 530 and 680 bookings. Using our model, we can write down the regression equations:*

$$\ln(\textit{bookings}) = -\alpha \cdot \textit{price} + \beta \cdot \textit{stars} + f_{hotel} + \epsilon \tag{8}$$

Here, we divide the unobservable ξ into a fixed effect f that is common for the same hotel (effectively a dummy binary variable), and an i.i.d. random error term ϵ. Using OLS, we get $\alpha = 0.0067$ and $\beta = 0.64$ which express the sensitivity of the consumers to price and their preference for "stars," respectively.

Of course, the assumption of homogeneity of consumer preferences is only an ideal case. In reality, consumers are different and their tastes vary. Next, we examine the case where the consumer have heterogeneous tastes.

3.2 Heterogeneous Consumers: BLP Model

In reality, consumers' preferences are heterogeneous. In principle, we could observe a customer for a long period of time and then use the Logit scheme described above to extract the preferences of each customer. Unfortunately, we can rarely observe individual behavior over long periods of time, so it is difficult to estimate the individual preferences for *each* consumer.

To allow preferences to vary, though, we can assume that preferences are a function of consumer demographics and purchase context. For example, everything else being equal, honeymooners may appreciate a hotel in a romantic remote setting, while business travelers may appreciate more a location with easy access to public transportation. We can therefore characterize each customer by a set of demographic characteristics (e.g., age, gender, travel purpose, etc.) and make the preference coefficients β to be a function of these demographics.

In this case, the overall preference distribution of the whole population is a *mixture* of preference distribution of the various consumer types in the population. The main challenge in this setting is that we only observe overall demand, and not the demand from each separate consumer group. So, the question

[3]Market share is defined as the percentage of total sales volume in a market captured by a brand, product, or firm.

[4]Since d_{total}, appears as a constant in across all equations, the absolute value of d_{total} and of the "buy nothing" demand d_0 is not relevant to the the parameter estimation and can be ignored.

becomes: *How can we find the preferences of various consumer types by simply observing the aggregate product demand?*

3.2.1 Model Specification

We solve this issue by monitoring demand for *similar* products in *different* markets, for which we know the *distribution* of consumers. Since the same product will have the same demand from a given demographic group, any differences in demand across markets can be attributed to the different demographics. The following simplified example illustrates the intuition behind this approach.

EXAMPLE 3. *Consider an example where we have two cities, A and B and two types of consumers: business trip travelers and family trip travelers. City A is a business destination with 80% of the travelers being business travelers and 20% families. City B is mainly a family destination with 10% business travelers and 90% family travelers. In city A, we have two hotels: Hilton (A_1) and Doubletree (A_2). In city B, we have again two hotels: Hilton (B_1) and Doubletree (B_2). Hilton hotels (A_1, B_1) have a conference center but no pool, and Doubletree hotels (A_2, B_2) have a pool but no conference center. To keep the example simple, we assume that preferences of consumers do not change when they travel in different cities and that prices are the same.*

By observing demand, we see that demand in city A (business destination) is 820 bookings per day for Hilton and 120 bookings for Doubletree. In city B (family destination) the demand is 540 bookings per day for Hilton and 460 bookings for Doubletree. Since the hotels are identical in the two cities, the changes in demand must be the result of different traveler demographics, hinting that a conference center is desirable for business travelers.

For this paper, to extract consumer preferences, we use the Random-Coefficient Model [6], introduced by Berry, Levinsohn, and Pakes, and commonly referred to as the *BLP model*. This model extends the basic Logit model by assuming the coefficients β and α in Equation 6 to be *demographic-specific*. Let T^i be a vector representing *consumer type*, which can specify a particular purchase context, age group, and so on. In the simplest case, we can have a binary variable for each consumer group. With the preferences being now demographic-specific, we write the utility surplus for consumer i, of type T^i, when buying product j, with features $\langle x_j^1, \ldots, x_j^k \rangle$, at price p_j to be:

$$US_j^i = \sum_k \beta^k(T^i) \cdot x_j^k - \alpha(I^i) \cdot p_j + \xi_j + \varepsilon_j^i. \qquad (9)$$

For the Logit model, in Equation 4, we used $V(\alpha, \beta)$ to stylistically separate the population preferences from the idiosyncratic behavior of the consumer. We now do the same for the *BLP* model, separating the mean population preferences from the demographic-specific preferences. So, we write $\beta^k(T^i) = \left(\bar{\beta}^k + \beta_T^k T^i\right)$, where $\bar{\beta}^k$ is the mean of the preference distribution, and β_T^k is a *vector* capturing the variation in the preferences from different consumer types. Similarly, we model α^i as a function of income I^i: $\alpha(I^i) = \left(\bar{\alpha} + \alpha_I I^i\right)$. Notice that we assume α_I and β_T to be independent. We rewrite US_j^i as:

$$US_j^i = \sum_k \left(\bar{\beta}^k + \beta_T^k T^i\right) \cdot x_j^k + \xi_j - \left(\bar{\alpha} + \alpha_I I^i\right) \cdot p_j + \varepsilon_j^i. \qquad (10)$$

We use $\delta_j = -\bar{\alpha} \cdot p_j + \sum_k \bar{\beta}^k \cdot x_j^k + \xi_j$ to represent the *mean* utility of product j. Then, as in the Logit model, we derive the choice probability for j, by integrating over the population demographic and income distributions $P(T)$ and $P(I)$:

$$P(choice_j) = \int \frac{\exp\left(\delta_j + \alpha_I I^i p_j + \sum_k \beta_T^k T^i x_j^k\right)}{1 + \sum_l \exp\left(\delta_l + \alpha_I I^i p_l + \sum_k \beta_T^k T^i x_l^k\right)} dP\,(T)\, dP\,(I) \qquad (11)$$

We explain next how we compute this integral and how we extract the parameters that capture the population preferences.

3.2.2 Estimation Methodology

With the model in hand, now we discuss how we identify the unknown parameters δ_j, α_I and β_T. We apply methods similar to those used in [6, 7] and [25]. In general, we estimated the parameters by searching the parameter space in an iterative manner, using the following steps:

1. Initialize the parameters $\delta_j^{(0)}$ and $\theta^{(0)} = (\alpha_I^{(0)}, \beta_T^{(0)})$ using a random choice of values.
2. Estimate market shares s_j given θ and δ.
3. Estimate most likely mean utility δ_j given the market shares.
4. Find the best parameters $\bar{\alpha}$ and $\bar{\beta}^k$ that minimize the unexplained remaining error in δ_j and evaluate the generalized method of moments (GMM) objective function.
5. Use Nelder-Mead Simplex algorithm to update the parameter values for $\theta = (\alpha_I, \beta_T)$ and go to Step 2, until minimizing the GMM objective function.

We describe the steps in more detail below.

Calculating market share s_j: To form the market equations (i.e., model predicted market share = observed market share), we need two things: the right-hand side s_j^{obs} that can be observed from our transaction data, and the left-hand side s_j, derived from Equation 11. Unfortunately, the integral in Equation 11 is not analytic. To approximate this integral, we proceed as follows: Given the demographic distribution, we "generate" a consumer randomly, with a known demographic and income and, therefore, known preferences. Then, using the standard Logit model (Equation 6), we generate the choice of the product for this consumer. For example, assume that we have the following joint demographic distribution of travel purpose and age group:

$$\begin{bmatrix} & Age \leq 45 & Age > 45 \\ Business & 15\% & 15\% \\ Family & 30\% & 40\% \end{bmatrix}.$$

In this case, we have a 40% probability of generating a "sample consumer" with family travel purpose and age above 45. By repeating the process and obtaining N_T samples of demographics T^i and N_I samples of income I^i, we can compute an unbiased estimator of the Equation 11 integral:[5]

$$s_j(\delta_j|\theta) \sim \frac{1}{N_I}\frac{1}{N_T}\sum_{I^i}^{N_I}\sum_{T^i}^{N_T} \frac{\exp\left(\delta_j + \alpha_I I^i p_j + \sum_k \beta_T^k T^{ik} x_j^k\right)}{1 + \sum_l \exp\left(\delta_l + \alpha_I I^i p_l + \sum_k \beta_T^k T^{ik} x_l^k\right)}. \qquad (12)$$

Estimate mean utility δ_j: Since we know how to compute market shares from the parameters, we can now find a value of δ_j that best "fits" the observed market shares. (Notice that, conditional on $\theta = (\alpha_I, \beta_T)$, market share s_j can be viewed as a function of the mean utility δ_j.) We apply the contraction mapping method recommended by Berry [6], which suggests computing the value for δ using an iterative approach:

$$\delta_j^{(t+1)} = \delta_j^{(t)} + (\ln(s_j^{obs}) - \ln(s_j(\delta_j^{(t)}|\theta))). \qquad (13)$$

The procedure is guaranteed to converge [6] and find δ_j that satisfies $s_j(\delta_j|\theta) = s_j^{obs}$.

[5] We use $N_T = N_I = 100$ in our study.

Minimize error, evaluate GMM objective function: Once we have the market shares and the mean utility parameters, we need to find the most likely demographic-specific weight deviations $\theta = (\alpha_I, \beta_T)$. Of course, different values for $\theta = (\alpha_I, \beta_T)$ will lead to different mean utilities and market shares. Hence, we need to find a criterion for identifying the best solution. We perform this in two steps: First, we use *Instrumental Variables (IV)* [13] to estimate the mean weights $\bar{\alpha}$ and $\bar{\beta}$, and extract the unobserved error term ξ from the mean utility function:

$$\xi(\theta) = \delta(\theta) - (\sum_k \bar{\beta}^k \cdot x^k - \bar{\alpha} \cdot p). \quad (14)$$

In our study, we use the average price of the "same-star rating" hotels in other markets as the instrument for price of a particular hotel to ensure that we do not have a correlation of the error term with a variable in our regression. Then, using the *generalized method of moments*,[6] we base our analysis on the moment condition that the mean of the unobserved error term ξ is *uncorrelated* with the instrumental variable IV. Thus, in our case we are trying to minimize the objection function:

$$GMMobj(\theta) = E[\xi'(\theta) \cdot IV]. \quad (15)$$

Iterate until GMM objective function is minimized: Once we identify the mean utility for a given set of weight deviations $\theta = (\alpha_I, \beta_T)$, we note the value of the GMM objective function $GMMobj(\theta)$. Then, we use the Nelder-Mead Simplex algorithm [21] to search for the optimal $\theta^* = (\alpha_I^*, \beta_T^*)$ that minimizes the GMM objective function.[7] This whole process eventually[8] identifies the heterogeneous weights that different consumers assign to product price, $\alpha(I^i) = \bar{\alpha}^* + \alpha_I^* \cdot I^i$, and those being assigned to product characteristics, $\beta(T^i) = \bar{\beta}^* + \beta_T^* \cdot T^i$.

EXAMPLE 4. *To illustrate this better, let's again look at Example 3, We know that, for business traveler, the utility surplus from hotel A_1 (conference center, no pool) is $US^B(A_1) = \delta_{A_1} + (\beta^B_{conf} \cdot 1 + \beta^B_{pool} \cdot 0) + \epsilon$, and for family travelers, the corresponding utility surplus is $US^F(A_1) = \delta_{A_1} + (\beta^F_{conf} \cdot 1 + \beta^F_{pool} \cdot 0) + \epsilon$. By $\beta^B_\bullet$ we denote the deviations from the population mean for business travelers towards "conference center" and "pool" and by $\beta^F_\bullet$ we denote the respective deviations for family travelers. Similarly, we can write down the utilities for hotels A_2, B_1 and B_2. Following the estimation steps discussed above, we discover that family travelers have $\beta^F_{conf} = \beta^F_{pool} = 0.5$. In other words, they have the same preferences regarding a pool and conference center. On the other hand, for business travelers, their preference towards "conference center" is much higher than towards "pool," with $\beta^P_{conf} = 0.9$ and $\beta^F_{pool} = 0.1$, respectively.*

Next, we explain how we leverage the above knowledge for building a better ranking model for product search.

4. RANKING USING UTILITY SURPLUS

So far, we have described models for inferring the preferences of consumers using a utility model and aggregate demand data. These models use the concept of surplus mainly as a conceptual tool to infer consumer preferences towards different product characteristics. In our work, the concept of surplus is directly used to find the product that is the "*best value for money*" for a given consumer. This simple idea is at the core of our work and as we will demonstrate in the experimental evaluation, it can lead to significant improvements in the quality of product search results.

Surplus-based Ranking: The first approach is to use the estimated surplus for each product and rank the available products in decreasing order of surplus. Therefore, at the top we will have the products that are the "best value" for consumers, for a given price. We define *Consumer Surplus* for consumer i from product j as the "normalized utility surplus," the surplus $\bar{US}_j^{(i)}$ divided by the mean marginal utility of money $\bar{\alpha}$.

$$CS_j = Normalized\_US_j = \sum_t \frac{1}{\bar{\alpha}} \bar{US}_j^{(i)}. \quad (16)$$

In the general, non-personalized case, if we were ranking products based on the "training" demand data then, in theory, our product ranking would be similar to a "best selling" ranking: The products that generate that largest surplus are the ones that would also generate the highest sales. (Notice that rational consumers prefer the products that generate the highest surplus.) However, when ranking products that are available *today*, the surplus-based ranking may be different for a variety of reasons: the *product price may have changed*, making some products a better "value for money," we may have a *new product* in the market, or the value of some product features may be time-dependent (e.g., the value of being next to a lake may be positive during warm weather and negative during the winter).

Personalized Surplus-based Ranking: In Section 3.2 we described how to estimate the value that consumers place on different product features, based on their own demographics and purchase context. The main outcome is that the value (surplus) that consumers get from a particular purchase is *different* than the average surplus for the overall population. This means that even the best possible ranking for the general population may not be optimal for an individual consumer.

Therefore, we extend our ranking to include a personalization component. To compute the personalized surplus, we can ask the consumer to give the appropriate demographic characteristics and purchase context (e.g., 35-49 years old, male, \$100K income, business traveler) and then use the corresponding deviation matrices β_T and α_I. It is then easy to compute the personalized "value for money" for this individual consumer, and rank products accordingly. Notice that the consumer *has the incentive to reveal demographics* in this scenario.

EXAMPLE 5. *For better understanding, let's re-consider the previous setting of the two hotels A_1 and A_2 for city A from Examples 3 and 4. Suppose that two consumers are traveling to city A on the same day: C_1, **a 35-49 years old business traveler, with an income \$50,000-100,000**, and C_2, **a 25-34 years old family traveler, with an income less than \$50,000**. Since these two travelers belong to different demographic groups and travel with different purposes, their preferences towards "conference center" and "pool" are different. Thus, the surplus they obtain from A_1 and A_2 varies. For example, the business traveler gets higher utility from A_1 due to the specialized conference center services, whereas the family traveler find A_2 more valuable due to the pool and price. This personalization component allows each consumer to identify the product that is the "best value for the money."*

---

[6] Due to space restrictions, we do not describe the GMM method in detail here. We refer the interested reader to [12] for further explanations.

[7] This approach is typically better than just following the function gradient. See `http://linux.math.tifr.res.in/programming-doc/gsl/gsl-ref_34.html` for an open source implementation and details.

[8] In our application, the computational time for each call (i.e., the inner loop) to the GMM objective function to solve for the mean utility is around 3 minutes on average. The global parameter search (i.e., the outer loop) by minimizing the GMM objective function takes an average of 20 calls. The total time for the estimation is around 40-60 minutes.

5. EXPERIMENTAL SETUP

For our experimental evaluation, we instantiated our model framework using as target application the area of *hotel search*.

Demand data: Travelocity, a large hotel booking system, provided us with the set of all hotel booking transactions, for 2117 randomly selected hotels over the United States. The transactions covered the period from November 2008 until January 2009. Based on the given transactions, we were able to compute the market shares of each hotel in each local market (i.e., metropolitan area), for each day.

Consumer demographics data: To measure the demographics of consumers in each target market, we used data provided by TripAdvisor: The consumers that write reviews about hotels on TripAdvisor also identify their *travel purpose* (*business, romance, family, friend, other*) and *age group* (*13-17, 18-24, 25-34, 35-49, 50-64, 65+*).[9] Based on the data, we were able to identify the types of travelers for each destination. To ensure the quality of the data, we computed the Jensen-Shannon divergence of the demographic distribution extracted from TripAdvisor with the corresponding traveler information from Travelocity, whenever available. The distributions were very similar with an average divergence of just 0.03.

Hotel location characteristics: We used geo-mapping search tools (in particular the Bing Maps API) and social geo-tags (from geonames.org) to identify the "external amenities" (e.g., shops, bars, etc) and public transportation in the area around the hotel. We also used image classification together with Mechanical Turk to examine whether there is a nearby beach, a nearby lake, a downtown area, and whether the hotel is close to a highway [16]. We extracted these characteristics within an area of 0.25-mile, 0.5 mile, 1-mile, and 2-mile radius.

Hotel service characteristics: We extracted the service characteristics from the reviews from TripAdvisor. Each review provides a general rating of the hotel, plus provides seven individual ratings on the following service characteristics: *Value, Room, Location, Cleanliness, Service, Check-in, and Business Service*. We used the average ratings of each hotel across these seven characteristics, together with the general review rating.[10] We also used the hotel description information from Travelocity, Orbitz, and Expedia, to identify the "internal amenities" (e.g., pool, spa, etc.)

Stylistic characteristics of online reviews: Finally, we extracted indicators that measure not the polarity of the reviews but rather some stylistic characteristics of the available reviews. We examined 2 text-style features: "subjectivity" and "readability" of reviews [10]. Also, since prior research suggested that disclosure of identity information is associated with changes in subsequent online product sales [9], we measured the percentage of reviewers for each hotel who reveal their real name or location information on their profile web pages.

Figure 2 shows the histograms of the important continuous variables, together with their correlations and scatterplots that illustrate the joint distributions of the variable pairs.

6. EXPERIMENTAL RESULTS

In the previous section, we have discussed how we retrieved different hotel characteristics through various sources. In this section, we present our findings from the empirical estimation. First, in Section 6.1 we present the results on estimating the model parameters, which correspond to the consumer preferences. Then, in Section 6.2, we show that our models generate significantly better rankings than the existing baselines.

Variable	Coef.	Variable	Coef.
Price	0.1157	# of competitors	-0.0930
Avg review length	0.0291	Crime	-0.4226
# of ext. amenities	0.0013	# of int. amenities	0.0048
Readability (SMOG)	0.2308	Beach	0.5498
Spelling errors	-0.0764	Lake	-0.1884
Avg. Subjectivity	-1.3468	Transport	0.00005
Dev. Subjectivity	-2.9106	Highway	0.2082
% of non-anon reviews	0.0892	Downtown	0.0161
Hotel Class	0.0317	# of reviews	-0.3897
Review Rating	0.0835	Review_Value	-0.2988
# of rooms	0.1525	Review_Location	0.0845
Review_Clean	0.1309	Review_Service	0.0105
Review_Checkin	-0.1151	Review_Bus_Service	0.1432

Table 1: Estimation results for mean weights (listing only statistically significant coefficients, with $p < 0.05$)

6.1 Interpretation of Estimated Weights

We present only the estimation results for the *BLP* model, as the results are strongly superior and closer to reality compared to the results from the *Logit* model.

Results for general population: Table 1 shows the estimation results for the mean weights, $\bar{\alpha}$ and $\bar{\beta}$, for the different variables in our model that have a statistically significant effect on demand. From the results, "beach" presents the highest positive impact compared to the other location characteristics. We also found significant impacts from service characteristics and quality characteristics of word-of-mouth. Meanwhile, "price" presents a positive sign, which is consistent with the "law of demand" in reality and indicates that the higher the price, the lower the demand. The negative sign on subjectivity means that customers are positive influenced by reviews that describe factual characteristics of hotels, and do not want to read personal stories of reviewers. (Notice that this is independent of the review polarity.) Notable is the negative sign on the *review_value* rating, which indicates that hotels that receive a "high value" rating have lower demand. This is *not* surprising: these are the hotels that are "undiscovered" and therefore have lower demand and prices than otherwise expected.

Results for specific demographics: We also obtained the demographic-specific deviations from the mean. We used *purchase context* and *age group* as the demographic dimensions for our experiments (see Section 5).

The value of demographic-specific deviation shows the "sensitivity of evaluation" for a product characteristic, within a particular consumer type. For example, customers on a romantic trip are more sensitive to hotel characteristics like "class" or "close to a beach", but they are less interested to know whether or not the hotel is close to highway exits. On the contrary, customers on a business trip are more sensitive to hotel characteristics like "internal amenities" or "easy access to highway", whereas they are likely influenced by the star rating of the hotel, compared to romance travelers. Figure 3 shows in details the evaluation deviations among different types.

We also examined age-specific preferences. Again, we found strong evidence for the deviation of weights associated with different age groups. Especially for "reviewer overall rating" and "review count", the deviations become quite striking. Figure 4, shows that customers from age 18 to 34 tend to be more sensitive to online reviews, compared to older ages. In particular, they

[9] There are other demographics available as well, such as "gender," "traveler residence location," "traveling with," "usual travel style" and "usual travel purpose." However, these fields were relatively sparsely populated. Therefore, we did not use these variables for our study.

[10] We have also extracted service-related variables by mining directly the text of the reviews [2, 23] but the additional information did not improve our model in a statistically significant manner.

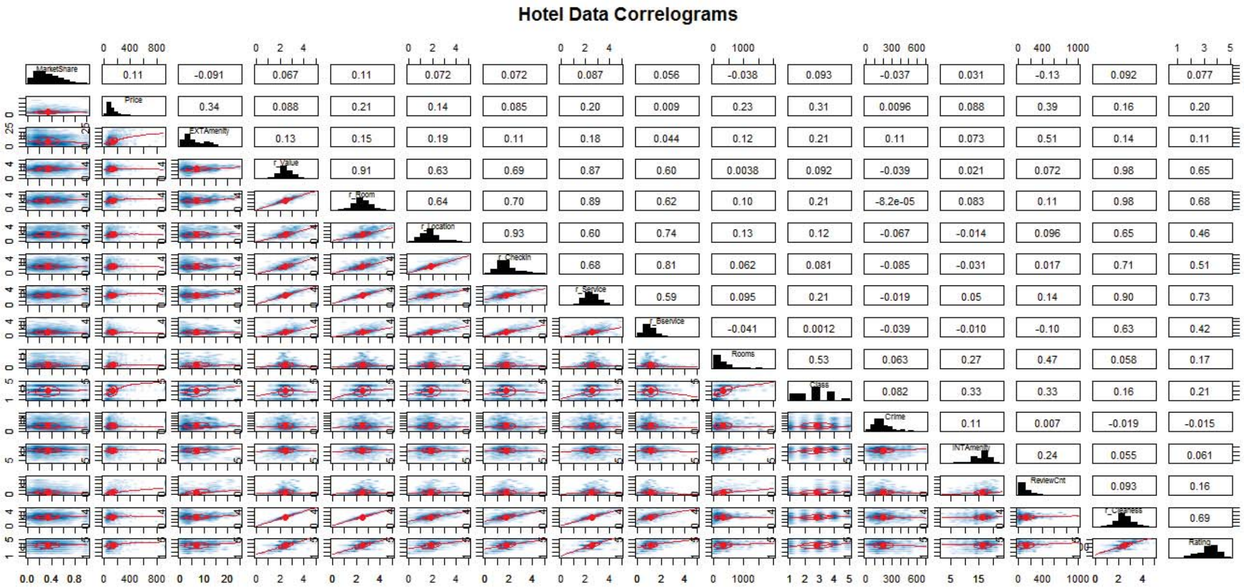

Figure 2: Diagonal: Histograms for the continuous variables in the data set (market share, price, ext_amenities, value, room, location, checkin, service, business service, rooms, class, crime, int_amenities, review_count, cleaningness, tripadvisor rating). Top: correlations of variable pairs. Bottom: scatterplots with ellipses for joint distributions of variable pairs.

City	TripAdvisor	Travelocity	Most Booked	Price Low to High	Price High to Low	Hotel Class	# of Reviews	# of Rooms	# of Amenities	Diversity: Price with Class
New York	**77%**	**63%**	**61%**	57%	**71%**	**88%**	**76%**	**89%**	*60%*	**80%**
Los Angeles	**72%**	58%	**71%**	*59%*	**84%**	**89%**	**87%**	**86%**	**69%**	**76%**
San Francisco	**79%**	57%	**65%**	**62%**	**70%**	**82%**	**68%**	**79%**	**79%**	**72%**
Orlando	**83%**	**81%**	**62%**	**63%**	**73%**	**79%**	**73%**	**79%**	**61%**	**79%**
New Orleans	**61%**	**69%**	*60%*	**78%**	**69%**	**80%**	**72%**	**91%**	58%	**85%**
Salt Lake City	**61%**	**80%**	**69%**	**66%**	**79%**	**83%**	**73%**	**70%**	**76%**	**79%**
Significance Level					$p = 0.05$ $\geq 56\%$	*$p = 0.01$ $\geq 59\%$*	**$p = 0.001$ $\geq 61\%$**		(Sign Test $N = 200$)	

Table 2: Consumer-surplus-based ranking vs. existing baselines. The percentage shows the number of users than indicated preference for the ranking of our surplus-based ranking, compared to the corresponding baseline ranking. The comparison was a pairwise blind comparison of rankings of 10 hotels. Each cell corresponds to a separate test using $N = 200$ users, for a total of 12000 comparisons. Across all comparisons, our ranking fares better in a statistically significant manner. Given the noise inherent in the Mechanical Turk ratings (some users may give random answers in such tests and vote equally for both approaches), the results show a strong superiority of the surplus-based rankings.

are 10-15 times more sensitive to online reviews compared to the age demographic of "65+ year old."

6.2 User Study: Ranking Comparison

With the estimated weights, we can derive the consumer surplus for each product (hotel), which can then be used to generate rankings. In our experiments, we used six metropolitan areas for which we generated hotel rankings (we used big cities, so that we have a meaningful number of hotels to rank).

Rankings based on the average consumer surplus vs. existing ranking baselines.: We used 10 different, existing rankings as a baseline for comparison, and we compared each baseline against our own surplus-based ranking. For each city and each ranking baseline, we performed pair-wise blind tests, asking 200 anonymous users on Amazon Mechanical Turk[11] to compare pairs of rankings and tell us which of the hotel ranking lists they prefer better. In all 60 comparisons, each using 200 users, the majority of users preferred our surplus-based rankings, in a statistically significant manner. Given that some MTurk users may be giving us random results, we expect the "real" performance of our algorithm to be even better than indicated in the numbers in Table 2.

Qualitative analysis: We also asked consumers *why* they chose a particular ranking, to better understand how users interpret the surplus-based ranking. Many users that liked our ranking indicated that our ranking promotes the idea that price is not the only main factor in rating the quality of products. Moreover, users strongly preferred the *diversity* provided by our ranking across both price and quality. In contrast, the other ranking approaches tend to list products of only one type (e.g., hotels with high review ratings are often very expensive hotels and the "most booked" are often 3-star mediocre hotels). We should emphasize at this point that our algorithm does not try explicitly to introduce diversity in the results. This is a direct outcome of our economic-based approach: If a segment of the market is systematically underpriced (hence making the "best deals" a homogeneous list), then we expect the market forces to fix this

[11] We restricted participation to US-based users. Mechanical Turk users are representative of the general US population.

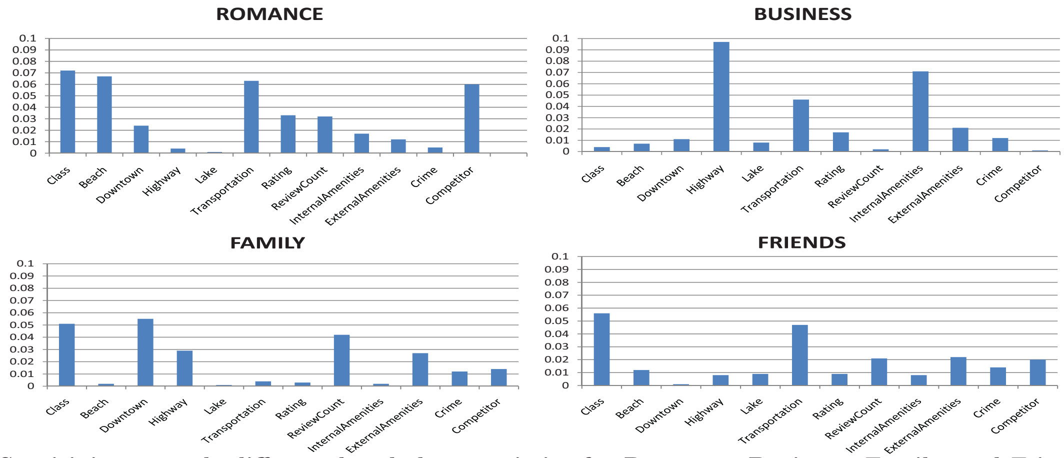

Figure 3: Sensitivity towards different hotel characteristics for Romance, Business, Family, and Friends-getaway travelers. We present absolute values, without polarity, to illustrate sensitivity on different product characteristics.

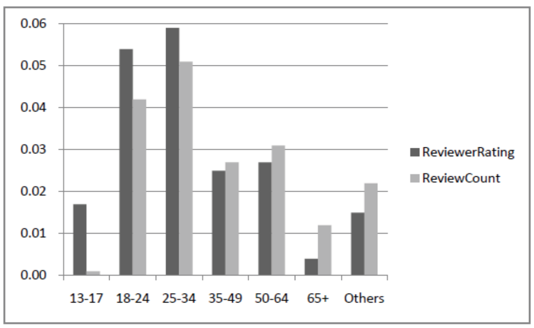

Figure 4: Sensitivity of different age groups to review rating & review count: Ages groups 18-24 and 25-34 pay more attention to online reviews compared to other age groups.

City	Business	Romance	Family	Friends
New York	77%	67%	81%	80%
Los Angeles	70%	65%	78%	69%
D.C.	79%	63%	70%	75%
Orlando	87%	85%	91%	84%

Table 3: Personalized surplus-based rankings vs. non-personalized, surplus-based rankings. The percentage shows the number of users than indicated preference for the personalized ranking compared to the general, non-personalized surplus-based ranking. Each cell of the table corresponds to $N = 200$ users, for a total of 3200 comparisons. In all cases, the personalized approach is better at the $p = 0.1\%$ level (sign test).

irregularity by increasing demand (and hence prices) for these products.

To test if we can generate same satisfaction levels by introducing diversity artificially, we also generated an additional "diversity ranking" using combined criteria of "price" and "hotel class." We interlaced the top-5 hotels with "the lowest price" and the top-5 hotels with "the highest ratings." The comparison of this "diversity-enabled" ranking against our algorithm also indicated that surplus-based diversity is better than an artificially-introduced diversity metric: In all cases, users strongly preferred the surplus-based ranking, as shown in the last column of Table 2.

Personalized rankings vs. rankings using average consumer surplus: We generated a few personalized rankings for different cities based on consumer-specific attributes, such as travel purpose. We conducted blind comparisons in a pair-wise fashion, based on 200 anonymous AMT users, for each comparison. Since we did not know the demographics of the users, we asked the MTurk workers to "select the hotel ranking you would prefer to use while trying to help a *business traveler* to book a hotel in *New York*"[12].

Based on the user responses, customers strongly preferred the personalized ranking that was tailored for a particular travel purpose using our technique ($p = 0.01$, sign test). For example, in our NYC experiment, 77% customers indicated their preferences towards the business-oriented ranking (ranking tailored for business travelers) rather than the average-level ranking, and 81% customers did so towards the family-oriented ranking (ranking tailored for family trip travelers). The results were similar across all tests.

Qualitative analysis: When we asked users' opinions for these comparisons, users did not bring up the issue of diversity. This was expected, as even our non-personalized rankings were already diverse. Instead, we found customers considering the context and expectations for a given trip. For example, users indicated that hotels for business trips do not necessarily need to be luxury, but need to provide a quiet business environment and easy access to highway or public transportation. On the other hand, for romantic trips, users strongly preferred the recommendations for luxury services with higher class rating. These results highly dovetail with our empirical estimation, suggesting that our ranking model indeed captures consumers' real purchase motivation behind the scene.

7. RELATED WORK

Our research is related to the work in online recommender systems, in particular, the content-based systems that recommend items similar to those that a user liked in the past (e.g., [20]). Content-based systems learn user preferences from the individual-level profiles elicited from users explicitly (e.g., through questionnaires or identified transactional behaviors). To compute a "content-based weight vector", a variety of techniques were used, such as the Rocchio algorithm, Bayesian classifiers, and Winnow algorithm [1]. Our research also leverages work on

[12] Of course, we substituted "business traveler' and "New York" with the appropriate values for each comparison.

learning consumer opinions from online reviews [23, 26]. In our work, through user modeling, we identify how users *behave* as a response to online reviews [2], and demonstrate how to extract demographic-specific preferences. Other studies proposed to combine popularity with user feedback or social annotations to refine search results [4, 14].

8. DISCUSSION AND FUTURE WORK

We presented a ranking algorithm that uses a behavioral model of consumers, based on utility maximization. The model generates an estimate of how much each product characteristic contributes to the product's overall utility, and estimates the sensitivity of consumers to changes in various product characteristics. The estimation models are privacy-friendly as they do not require individual consumer data but rather rely on aggregate data. Based on the generated models, we can estimate the surplus that each product generates for each consumer, and build rankings that capture the user preferences. We demonstrated, through extensive user studies, that our ranking schemes are better than any of the existing baselines. We also showed that personalized surplus-based rankings are even better than the non-personalized surplus-based rankings. By doing so, we are able to target at each individual customer, and offer products with the "*best value for money*" in response to consumer queries.

We should also note that our ranking scheme is "causal," in the sense that the model can predict what "should" happen when we observe changes in the market. For example, when we see a new product in the marketplace, we can rank it by simply observing its characteristics, without waiting to see the consumers' demand for the product. Also, we can dynamically change the rankings as a reaction to changes in the products. For example, if we observe a price change, or if we observe that a hotel closes its pool for renovations, we can adjust immediately the surplus values and re-estimate the rankings.

Also, in order to better understand the antecedents of consumer's decisions, future work can look not only at transaction data but also into their browsing history and learning behavior. For example, our current model assumes that consumers are engaging into optimal utility maximizing behavior. However, this is not always true, as some consumers are more thorough than others in their search. By leveraging browsing histories, we can build models that explicitly take into consideration the fact that some users are "utility optimizers" and some others simply engage into "satisficing." It would be also interested in examining the difference in the conversion rate of users, when presented with surplus-based rankings.

By examining product search through the "economic lens" of consumer behavior, we can leverage micro-economic theory and many theoretical models that have been developed over the years, which try to capture the decision-making process of humans. Economic theory provides a very solid basis upon which we can build further computer science research, which has a different focus than economic research. Our example is illustrating: while economists have been building utility models for years, their goal was to estimate demand for products and the notion of surplus was just "a means to an end" and never had of value by itself. By focusing on product ranking, we showed how surplus can improve product search. Our experimental results demonstrated a significant improvement in user satisfaction. Other economic models (e.g., measuring the utility of product bundles) can also be directly used in consumer-facing applications on the Web (e.g., search for "product bundles" instead of simple products). We are very optimistic that this interdisciplinary research direction can generate very interesting results in the future.

References

[1] Adomavicius, G., and Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE TKDE 17* (2005), 734–749.

[2] Archak, N., Ghose, A., and Ipeirotis, P. G. Show me the money!: deriving the pricing power of product features by mining consumer reviews. In *KDD* (2007), pp. 56–65.

[3] Balke, W.-T., and Güntzer, U. Multi-objective query processing for database systems. In *Proceedings of 28th International Conference on Very Large Data Bases (VLDB)* (2004), pp. 936–947.

[4] Bao, S., Wu, X., Fei, B., Xue, G., Su, Z., and Yu, Y. Optimizing web search using social annotations. In *WWW* (2007).

[5] Berry, S. Estimating discrete choice models of product differentiation. *RAND Journal of Economics 25* (1994), 242–262.

[6] Berry, S., Levinsohn, J., and Pakes, A. Automobile prices in market equilibrium. *Econometrica 63* (1995), 841–890.

[7] Berry, S., and Pakes, A. The pure characteristics demand model. *International Economic Review 48* (2007), 1193–1225.

[8] Chevalier, J. A., and Goolsbee, A. Measuring prices and price competition online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics 1*, 2 (2003), 203–222.

[9] Forman, C., Ghose, A., and Wiesenfeld, B. Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets. *ISR 19*, 3 (2008), 291–313.

[10] Ghose, A., , and Ipeirotis, P. G. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE TKDE* (2010).

[11] Ghose, A., Ipeirotis, P., and Sundararajan, A. Opinion mining using econometrics: A case study on reputation systems. In *ACL* (2007).

[12] Hansen, L. Large sample properties of generalized method of moments estimators. *Econometrica 50*, 4 (1982), 1029–1054.

[13] Heckman, J. Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. *Journal of Human Resources 32*, 3 (1997), 441–462.

[14] Jin, R., Valizadegan, H., and Li, H. Ranking refinement and its application to information retrieval. In *WWW* (2008).

[15] Lancaster, K. *Consumer Demand: A New Approach*. Columbia University Press, New York, 1971.

[16] Li, B., Ghose, A., and Ipeirotis, P. G. Stay elsewhere? improving local search for hotels using econometric modeling and image classification. In *WebDB* (2008).

[17] Marshall, A. *Principles of Economics*, Eighth ed. Macmillan and Co., London, 1926.

[18] McFadden, D. *Conditional Logit Analysis of Qualitative Choice Behavior*. Academic Press, New York, 1974.

[19] McFadden, D., and Train, K. Mixed MNL models of discrete response. *Journal of Applied Econometrics 15*, 5 (2000), 447–470.

[20] Mooney, R., and Roy, L. Content-based book recommending using learning for text categorization. In *ACM SIGIR Workshop Recommender Systems: Algorithms and Evaluation* (1999).

[21] Nelder, J. A., and Mead, R. A simplex method for function minimization. *The Computer Journal 7*, 4 (1965).

[22] Nie, Z., Wen, J.-R., and Ma, W.-Y. Webpage understanding: beyond page-level search. *SIGMOD Record 37*, 4 (2008), 48–54.

[23] Pang, B., and Lee, L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval 2*, 1-2 (2008).

[24] Rosen, S. Hedonic prices and implicit markets: Product differentiation in pure competition. *J. of Political Econ. 82*, 1 (1974), 34–55.

[25] Song, M. A hybrid discrete choice model of differentiated product demand with an application to personal computers. FR 08-09, 2008.

[26] Ye, Q., Law, R., and Gu, B. The impact of online user reviews on hotel room sales. *Int. J. of Hosp. Mgmnt. 28*, 1 (2009), 180–182.

[27] Yee, K.-P., Swearingen, K., Li, K., and Hearst, M. Faceted metadata for image search and browsing. In *CHI* (2003), pp. 401–408.