Improving Product Search with Economic Theory

Beibei Li

Supervised by Panagiotis G. Ipeirotis and Anindya Ghose

Department of Information, Operations & Management Sciences
Stern School of Business, New York University
44 West 4th street, New York, NY 10012, USA
bli@stern.nyu.edu

***Abstract*— With the growing pervasiveness of the Internet, online search for commercial goods and services is constantly increasing, as more and more people search and purchase goods from the Internet. Most of the current algorithms for product search are based on adaptations of theoretical models devised for "classic" information retrieval. However, the decision mechanism that underlies the process of *buying a product* is different than the process of *judging a document as relevant* or not. So, applying theories of relevance for the task of product search may not be the best approach.**

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 *consumer surplus* after the purchase. In a sense, we rank highest the products that are the "best value for money" for a specific user. Our approach naturally builds on decades of research in the field of economics and presents a solid theoretical foundation in which further research can build on. We instantiate our research by building a search engine for hotels, and show how we can build algorithms that naturally take into account consumer demographics, heterogeneity of consumer preferences, and also account for the varying price of the hotel rooms. Our extensive user studies demonstrate an overwhelming preference for the rankings generated by our techniques, compared to a large number of existing strong baselines.

I. INTRODUCTION

It is now widely acknowledged that online search for products is increasing in popularity, as more and more users search and purchase products from the Internet. Most of the current attempts for product search are based on models of relevance from "classic" information retrieval theory. However, the decision mechanism that underlies the process of *buying a product* is different from the process of *judging a document as relevant* or not. Customers try to identify the "best" deal that satisfies their specific desired criteria but without compromising on the price. Furthermore, 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, average customer rating. This approach has quite a few shortcomings. First, it ignores the *multidimensional* preferences of consumers. Second, it fails to leverage the *textual* information generated by the online communities beyond the numerical ratings. Third, it hardly takes into account the *heterogeneity* of consumers. These drawbacks highly necessitate a new recommendation strategy for product search that can better understand consumers' underlying purchase behavior, in order to capture their multidimensional preferences as well as heterogeneous tastes.

Some recent studies proposed to combine popularity with user feedback or social annotations to refine search results [1], [2]. One such application is recommender systems. Nevertheless, to the best of our knowledge, existing techniques in this field still have limitations. First, the current recommendation mechanisms require consumers' proactive efforts to log into the system. However, in reality many consumers browse anonymously and give their details just for purchasing. Consequently, recommendations would suffer from a "cold start" *every* time, even for a return customer. Second, consumers are reluctant to give out their individual demographic information due to privacy concerns. Therefore, most context information is missing at the individual consumer level. Third, for certain goods with a low purchase frequency for an individual consumer (such as hotel or real estate), there are hardly any repeated purchases one could leverage towards building a predictive model (i.e., models based on collaborative filtering).

In this thesis, we motivate our focus on these issues and propose to design a new ranking system for recommendation that leverages economics modeling. It aims at making recommendations based on better perception of the underlying *causality* of consumers' purchase decisions. 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 of each individual.

Our study is instantiated on a unique data set containing transactions from 2008/11 to 2009/1 for US hotels from a major travel website. The final outcome allows us to infer the *weights* that consumers assign to each individual product characteristic. Based on this, we derive the *consumer surplus* from each product, which represents how much extra value one can obtain by purchasing a product. Then we rank all the products accordingly. This ranking strategy is then extended to a personalized level based on the distribution of consumers' demographics. Such a personalized ranking can be recommended in response to a user query on search engines to assist the consumer in locating products with specified criteria and the "best value for money". In contrast to the existing research in recommender systems which tends to give recommendations using a machine learning-based "black-box" style, this economics-based approach tries to capture the overall decision-making process of consumers.

II. Theory Model

In this section, we provide the background from economic theory that explains the basic concepts behind our model. We start with the *expected utility* theory, *characteristics-based* theory, and *consumer surplus*. Then we discuss how we leverage these concepts in our setting and empirically estimate the parameters for our model, even in the absence of information about the characteristics of individual customers.

A. Choice Decisions and Utility Maximization

Our model is derived from economics, and in particular from *expected 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." The rationality assumption defines that each person tries to maximize its own utility.

More formally, assume that the consumer has a choice across products $X_1, \ldots, X_n$, and each product X_j has a price p_j. Buying a product involves the exchange of money for a product. Therefore, to analyze the purchasing behavior we need to have two components for the utility function:

- *Utility of Product:* The utility that the consumer will get by buying the product X_j.
- *Utility of Money:* The utility that the consumer will lose by paying the price p_j for product X_j.

On one hand, the decision to purchase product X_j generates a product utility $U(X_j)$. According to Lancaster's *characteristics theory* [3] and Rosen's *hedonic price model* [4], we assume that differentiated products are described by vectors of objectively measured characteristics. Let x_j^k denote the kth observed characteristics of product X_j. Thus, the utility of product can be defined as the aggregation of weighted utilities of observed individual characteristics and an unobserved characteristic, ξ_j, as follows

$$U(X_j) = U(x_j^1, \ldots, x_j^k) = \sum_k \beta_j^k \cdot x_j^k + \xi_j. \tag{1}$$

On the other hand, assume that the consumer has some disposable income I that generates a money utility $U(I)$. Paying the price p_j decreases the money utility to $U(I - p_j)$. We typically assume that p_j is relatively small compared to the disposable income I, and the *marginal utility* of money remains constant in the interval $I - p_j$ to I [5]. In this case,

$$U(I) - U(I - p_j) = \alpha I - \alpha(I - p_j) = \alpha p_j. \tag{2}$$

With the assumption of rationality, a consumer purchases product X_j if and only if it provides him with the highest increase in utility. Let *consumer surplus* denote the "increase" in utility after purchasing a product. This idea naturally generates a ranking order: The products that generate the highest consumer surplus should be ranked on top.

B. Consumer Heterogeneity

As we can see, the key for this model is to identify the different product characteristics and estimate the corresponding weights assigned by consumers towards the characteristics and the price of the product. However, different consumers may hold different evaluations towards the product characteristics and towards the money. In order to capture the consumer heterogeneity, we use the Random-Coefficient Logit Model from econometrics [6] (by Berry, Levinsohn and Pakes commonly referred to as BLP model).

This model incorporates consumer heterogeneity by assuming that consumers have idiosyncratic tastes towards product characteristics. In other words, the coefficients β and α in equation 1 and 2 are different for each consumer. Based on this, we define the excess utility for consumer i to buy product X_j as

$$\begin{aligned} U_j^i &= U_h(X_j) - [U_m(I^i) - U_m(I^i - p_j)] + \varepsilon_j^i \\ &= \underbrace{\sum_k \beta^{ik} \cdot x_j^k + \xi_j}_{\text{Utility of product}} - \underbrace{\alpha^i p_j}_{\text{Utility of money}} + \underbrace{\varepsilon_j^i}_{\text{Stochastic error}}. \end{aligned} \tag{3}$$

Here, I^i is the income of consumer i, p_j is the price of product X_j, U_m is the utility of money (parameterized by user specific weight scalar α^i), and U_h is the utility of product purchased (parameterized by user specific weight vector β^i). Note that ξ is a *product-specific* disturbance scalar summarizing unobserved characteristics of product X_j, whereas ε_j^i is a stochastic choice error term that is assumed to be i.i.d. across *products and consumers* in the selection process. The parameters to be estimated are α^i and β^i, which represent the weights that consumer i assigns towards "money" and towards different observed product characteristics, respectively.

To make this model tractable, we make some assumptions about α^i and β^i. We assume that these weights are distributed among consumers by some known statistical distribution, $\alpha^i \sim (\alpha^i | \bar{\alpha}, \delta_\alpha)$ and $\beta^i \sim (\beta^i | \bar{\beta}, \delta_\beta)$. Our goal is then to estimate the means $(\bar{\alpha}, \bar{\beta})$ and the standard deviations $(\delta_\alpha, \delta_\beta)$ of these two distributions.

Furthermore, we notice that these heterogeneities result from some particular demographic attributes of consumers. Therefore, we assume that $\delta_\alpha \sim \alpha_I I^i$, where I^i represents consumer income; $\delta_\beta \sim \beta_T T^i$, where T^i is a vector representing *consumer type*, which specifies a particular purchase context, age group, etc. So $\alpha^i = \left(\bar{\alpha} + \alpha_I I^i\right)$ and $\beta^i = \left(\bar{\beta} + \beta_T T^i\right)$. We then rewrite our model as follows

$$U_j^i = \left(\bar{\alpha} + \alpha_I I^i\right) \cdot p_j + \sum_k \left(\bar{\beta}^k + \beta_T^k T^{ik}\right) \cdot x_j^k + \xi_j + \varepsilon_j^i. \tag{4}$$

Let $\delta_j = -\bar{\alpha} p_j + \sum_k \bar{\beta}^k \cdot x_j^k + \xi_j$ represent the *mean* utility of product X_j, we can get the following:

$$U_j^i = \delta_j + \alpha_I I^i p_j + \sum_k \beta_T^k T^{ik} x_j^k + \varepsilon_j^i. \tag{5}$$

Following McFadden's Logit discrete choice model [7], the market share s_j of product X_j can be calculated as the proportion of consumers who choose product X_j over other products in the same market:

$$s_j = \int \frac{exp\left(\delta_j + \alpha_I I^i p_j + \sum_k \beta_T^k T^{ik} x_j^k\right)}{1 + \sum_l exp\left(\delta_l + \alpha_I I^i p_l + \sum_k \beta_T^k T^{ik} x_l^k\right)} dP\left(T\right) dP\left(I\right), \quad (6)$$

where $dP\left(\cdot\right)$ denotes population distribution functions.

C. Estimation Methodology

Recall that our goal here is to estimate the mean and variance of α_i and β_i. We apply estimation methods similar to those used in [6], [8]. This problem can be essentially reduced to a procedure of solving a system of nonlinear equations. We solve it using the *Generalized Method of Moments (GMM)* from econometrics [9]. In general, with a given starting value of $\theta_0 = (\alpha_I^{(0)}, \beta_T^{(0)})$, we look for the mean utility such that the model predicted market share from equation 6 equates the observed market share. From there, we form a GMM objective function using the moment conditions such that the mean of unobserved characteristics is uncorrelated with instrumental variables. Based on this, we identify a new value of $\theta_1 = (\alpha_I^{(1)}, \beta_T^{(1)})$, which is used as the starting point for the next round iteration. This procedure is repeated until the algorithm finds the optimal value of that minimizes the GMM objective function. To find a solution, we applied the contraction mapping method suggested by [6].

III. Data

We have complete information on all transactions conducted over a 3 month period from 2008/11 to 2009/1 for 2117 randomly selected hotels in the US. Further, we have data on hotel attributes from 4 sources: (i) location characteristics, (ii) service characteristics, (iii) review characteristics, and (iv) reviewer characteristics.

Location characteristics: We used geo-mapping search tools and social geo-tags to identify different "external amenities" (such as shopping malls, restaurants, etc). However, since some location-based characteristics, such as "near the beach" and "near downtown", are not directly measurable based on reviews, tags or any geo-mapping search services, we used image classification techniques to infer such features from the satellite images of the area. We extracted hybrid satellite images with 4 different zoom levels using the Visual Earth Tile System. Then we performed SVM classification. The results showed an accuracy of 91.2% for "Beach" feature and 80.7% for "Downtown" feature. Nevertheless, some characteristics are even harder to identify by image classification algorithms, such as "near the highway", we acquire them through human annotations using Amazon Mechanical Turk.

Service characteristics: With regard to the service-based hotel characteristics, we extracted them from the website of TripAdvisor. Since hotel amenities are not directly listed on TripAdvisor website, we retrieved them by following the link provided on the hotel web page, which directs to one of its cooperative partner websites (i.e., Travelocity, Orbitz, etc.).

Review characteristics: We examined 2 text-style features: "subjectivity" and "readability" of reviews [10]. To better capture the review text-style, we used a multiple-item method. We included 2 sub-features for subjectivity and 5 sub-features for readability, each of which measures the review text-style from an independent point of view. In order to decide the probability of subjectivity for review text, we trained a classifier using as "objective" documents the hotel descriptions of each of the hotels in our data set. We randomly retrieved 1000 reviews to construct the "subjective" examples in the training set. We conducted the training process by using a 4-gram Dynamic Language Model classifier provided by the LingPipe toolkit. Thus, we were able to acquire a subjectivity confidence score for each sentence in a review, thereby deriving the mean and standard deviation of this score, which represent the probability of the review being subjective.

Reviewer characteristics: Finally, previous research suggested that the prevalence of reviewer disclosure of identity information is associated with changes in subsequent online product sales [11]. Therefore, we decide to include one particular characteristic capturing the level of reviewers' disclosure of their identity information on these websites, "real name or location." These different data sources are then merged to create one comprehensive dataset.

IV. Ranking

After we have estimated the parameters in our model and inferred the economic values of product characteristics, we now discuss how to derive the consumer surplus from our model. From there, we propose a new ranking approach for products based on the consumer surplus.

A. Consumer Surplus-based Ranking

In general, we are interested to know what the excess utility, or consumer surplus, is for consumers on an average level to choose a product. Therefore, we define consumer i's surplus from product X_j as the sum of X_j's mean excess utility $\bar{U}_j^{(i)}$ divided by the mean price elasticity $\bar{\alpha}$ over all markets t.

$$CS_j = \sum_t \frac{1}{\bar{\alpha}} \bar{U}_{jt}^{(i)}. \quad (7)$$

We thereby propose a new ranking approach based on this. The basic idea is to rank products by their consumer surplus, i.e., how much "extra value" consumers can obtain after paying for the price. If a product provides a comparably higher surplus for consumers on average, then it should appear on the top level of the ranking list and should be highly recommended to consumers.

This model can be further extended to a personalized level since consumers often have idiosyncratic tastes towards product characteristics, hence obtaining different surplus even for the same product. Specifically, for a consumer of a particular demographic type, we derive a personalized ranking based on her individual consumer surplus. We achieve this by interpolating the estimated weight deviation matrix β_T and vector α_I into our model which capture the way that demographics influence the preferences of users. By doing so, we are able to compute the consumer-specific utility of a particular product for each consumer by interacting her own demographic attributes with the estimated weights. Based on this, a personalized consumer surplus can thereby be derived for each individual consumer. This personalization component allows us to further facilitate our consumer surplus-based ranking approach. It can help each individual consumer identify her own "best bang for the buck".

B. Evaluation With User Study

To evaluate the quality of our ranking technique, we conducted user study using Amazon Mechanical Turk (AMT).

First, we compared rankings based on the average consumer surplus with existing baselines. We generated different rankings for the top-10 hotels, according to 6 current existing baseline criteria: price low to high, price high to low, maximum online review count, hotel class, hotel size (number of rooms), and number of internal amenities. Then, we performed pair-wise blind tests, asking 100 anonymous AMT users to compare pairs of rankings and tell us which of the hotel ranking lists they prefer the most. We tested the results for a few large cities like New York city, and the results were highly encouraging. A large majority of customers preferred our ranking when listed side-by-side with the other competing baseline techniques ($p = 0.05$, sign test).

Furthermore, we compared personalized ranking with the average-level ranking. We generated a few personalized rankings for different cities based on consumer-specific attributes, such as travel purpose. Again, we conducted blind comparisons in a pair-wise fashion based on 100 anonymous AMT users. Based on the user responses, customers preferred the ranking that was tailored for a particular travel purpose using our technique ($p = 0.01$, sign test). For example, in our NYC experiment, 80% customers indicated their preferences towards the business-oriented ranking (ranking tailored for business travellers) rather than the average-level ranking, and 87% customers did so towards the family-oriented ranking (ranking tailored for family trip travellers). In the mean time, we also found similar trends in smaller cities. For instance, in our Orlando experiment, 91% and 95% customers chose the business- and family-oriented ranking over the average-level ranking, respectively.

We also asked consumers why they chose a particular ranking, to understand better how users interpret the surplus-based ranking. The majority of the users indicated that our consumer surplus-based ranking promotes the idea that price is not the only main factor in rating the quality of products. Instead, a good ranking recommendation should be able to satisfy multidimensional preferences. 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., very expensive ones). Based on the qualitative opinions of the users, it appears that diversity in product choices is indeed an important factor that improves the satisfaction of consumers, and an economic approach for ranking introduces diversity naturally. However, this effect may be less pronounced under a personalized ranking. In our second experiment, we found customers with personal contexts and demographic attributes tend to have more specific expectations towards the ranking recommendations. These user study results highly dovetail with our empirical estimation, which strongly suggests that our economic-based ranking model indeed captures consumers' real purchase motivation behind the scene.

V. Conclusions

In this thesis, we propose an economic theory-based model for product search. We rank highest the products that generate the highest *consumer surplus* after the purchase. Our user studies demonstrate an overwhelming preference for our ranking, compared to existing strong baselines. In the future, to better evaluate our techniques, we plan to obtain more individual level context information that is available during purchase. With richer individual level information, we are able to conduct traditional collaborative filtering or content-based algorithms, hence comparing our personalization results with theirs. Moreover, we also consider working with travel search companies to test our results when depolyed in a real environment. We have already started discussion along these lines with some companies who have expressed strong interest in collaborating with us.

References

[1] R. Jin, H. Valizadegan, and H. Li, "Ranking refinement and its application to information retrieval," in *WWW2008 Proceedings*, 2008, pp. 397–406.

[2] S. Bao, X. Wu, B. Fei, G. Xue, Z. Su, and Y. Yu, "Optimizing web search using social annotations," in *WWW2007 Proceedings*, 2007, pp. 501–510.

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

[4] S. Rosen, "Hedonic prices and implicit markets: Product differentiation in pure competition," *Journal of Political Economy*, vol. 82, no. 1, pp. 34–55, 1974.

[5] A. Marshall, *Principles of Economics*, 8th ed. London: Macmillan and Co., 1926.

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

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

[8] S. Berry and A. Pakes, "The pure characteristics demand model," *International Economic Review*, vol. 48, pp. 1193–1225, 2007.

[9] L. Hansen, "Large sample properties of generalized method of moments estimators," *Econometrica*, vol. 50, no. 4, pp. 1029–1054, 1982.

[10] A. Ghose and P. G. Ipeirotis, "Estimating the socio-economic impact of product reviews: Mining text and reviewer characteristics," 2008, working paper, New York University.

[11] C. Forman, A. Ghose, and B. Wiesenfeld, "Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets," *Information Systems Research*, vol. 19, no. 3, pp. 291–313, 2008.