

Towards a Theory Model for Product Search

Beibei Li, Anindya Ghose, Panagiotis G. Ipeirotis

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

WWW 2011, Hyderabad, India
3/31/2011

I Want to Buy...

“How can I find the best hotel in New York City?”

Great Price for what it offers!

Great Customer Reviews...

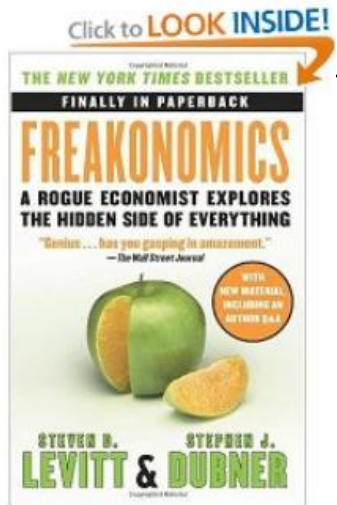
Great Location...

Great Service...



Ranking with single criterion is not enough!

Recommender Systems?



Customers Who Bought This Item Also Bought

Recommended for You
These recommendations are based on items you own and more.

view: **All** | [New Releases](#) | [Coming Soon](#)

1. [LOOK INSIDE!](#) [Econometric Analysis of Cross Section](#)

Problem:

- For many products, low purchase frequency;
- Privacy: *Individual-level* purchase history to derive personal preference.

Research Challenge and Goal

How to find the *best* product?

- Low purchase frequency
- Privacy
- Heterogeneity
- Multi-dimensional preferences

A New Ranking System for Product Search

- ✓ Economic utility theory, “Best Value” ranked on Top;
- ✓ Infer personal preferences from *aggregate & anonymous* data;
- ✓ Validated with user study with +15000 users, 6 cities.

Theoretical Background

How to define “Best Value”?



Get Hotel
(happy)

$>=?<$



Pay Money
(unhappy)

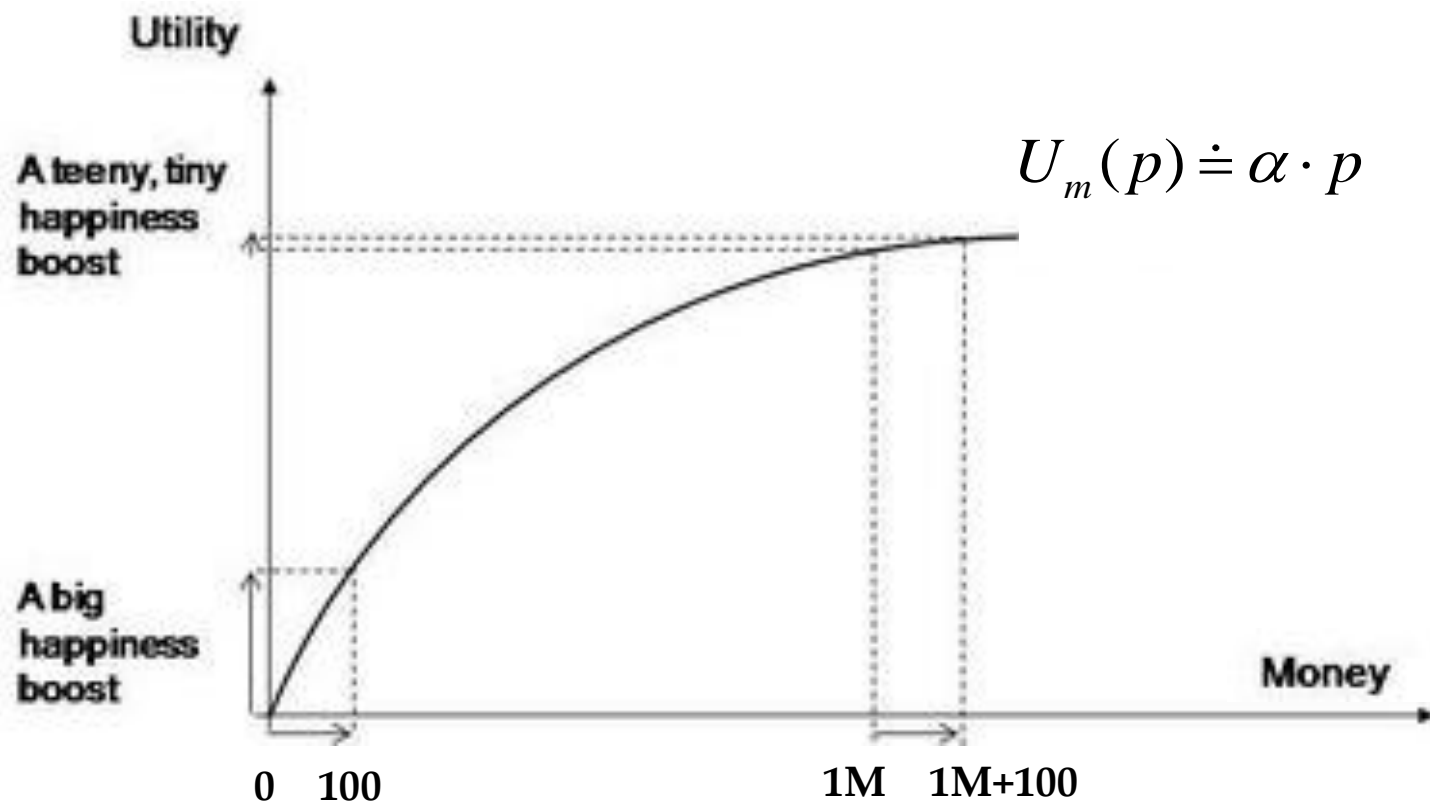
Utility: Quantify the happiness.

Utility of Products

Utility of Money

Utility of Money

- **Utility of Money** – The utility that the consumer will *lose* by paying the price for that product.



Utility of Product

- **Utility of Product** – The utility that the consumer will *gain* from buying the product.

Simplest case using a linear combination:

$$U_p(X) = U_p(x^1, \dots, x^K) = \sum^K \beta^k \cdot x^k + \xi.$$

Latent Consumer Preferences

Observed Product
 Characteristics

Utility Surplus

Utility Surplus for a consumer is the *gain* in the utility of product minus the *loss* in the utility of money.

$$\text{Utility Surplus} = \underbrace{\sum_{k=1}^K \beta^k \cdot x^k + \xi}_{\text{Utility of Product}} - \underbrace{\alpha \cdot p}_{\text{Utility of Money}} + \underbrace{\varepsilon}_{\text{Stochastic Error}}$$

- The higher the surplus, the higher “value” from a product.

Key Challenge: How to estimate the preferences?

Logit Model (McFadden 1974)

Assumption: Everyone likes things the same, but also has a personal component of choice, the error term ε .

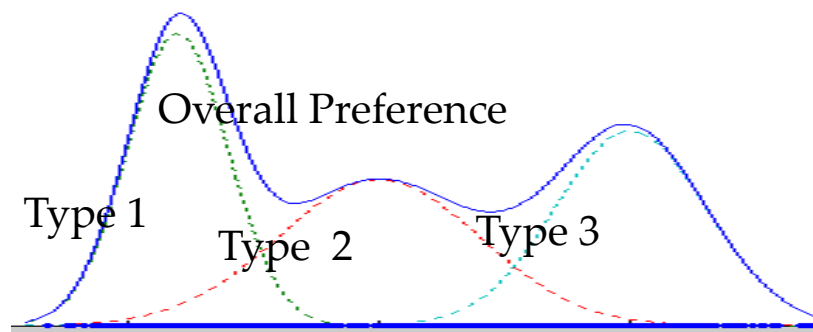
$$\begin{aligned}\text{Observed Market Share}_j &= \Pr(\text{Consumers choose } j \text{ over everything else}) \\ &= \Pr(\text{Surplus}_j > \text{Surplus_everything else})\end{aligned}$$

Solution by logistic regression!

Notice: Logistic Regression is a direct derivation from a theory-driven user behavior model. (McFadden 2000, Nobel Prize)

But, consumers are different...

BLP Model (Berry, Levinsohn, and Pakes 1995)



BLP Model:

- All consumers are not the same;
- Consumers belong to groups with different preferences;
- Group preference defined through consumer demographics, income, purchase purpose, ..., etc.

$T = [\text{age, gender, income, purpose, ...}] \rightarrow \text{Preference} = f(T)$

Problem:

We do *NOT* know T for each *individual* consumer at the purchase.

BLP Model (Berry, Levinsohn, and Pakes 1995)

What do we know?

- Demographic distributions!
- Demographic differences in different markets!
- Overall demand in different markets!

Basic Idea: Monitor demand for similar products in different markets.

differences in demand → different demographics

BLP Model - Example

Example 2: Lunch Buffet

Lamb Roganjosh: Stewed lamb with Kashmiri chillies;

Chicken Alfredo: Flat pasta cooked with cream and cheese;



Table A: 80% Indians, 20% Americans;

- Lamb: 80% gone, Chicken: 20% gone.

Table B: 10% Indians, 90% Americans;

- Lamb: 10% gone, Chicken: 90% gone.

→ Indians favor lamb, and Americans favor chicken!

BLP: Aggregate Demand → Individual Preference

Surplus-based Ranking

Basic Idea: Compute the surplus for each product based on the estimated consumer preferences, and rank the products accordingly.

- Top-ranked product provides “best value”!

Again, people are different...

Personalized Ranking: by incorporating consumer demographics and purchase context, we can further derive *personalized* surplus.

Hotel Search Experiment - Data

Demand Data: *Travelocity*, 2117 hotels, 11/2008-1/2009.

Demographics: *TripAdvisor*, “Travel purpose,” “Age group.”

Location Characteristics:

- Social geo-tags
- Geo-Mapping Search Tools
- Image Classification

Service Characteristics: *TripAdvisor & Travelocity*

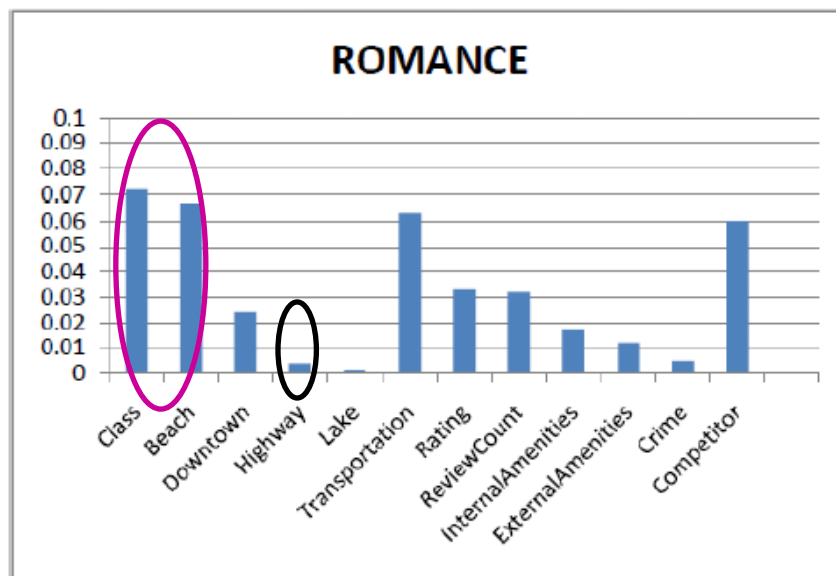
Stylistic Characteristics for the quality of word-of-mouth:

- Text Mining: “Subjectivity,” “Readability”

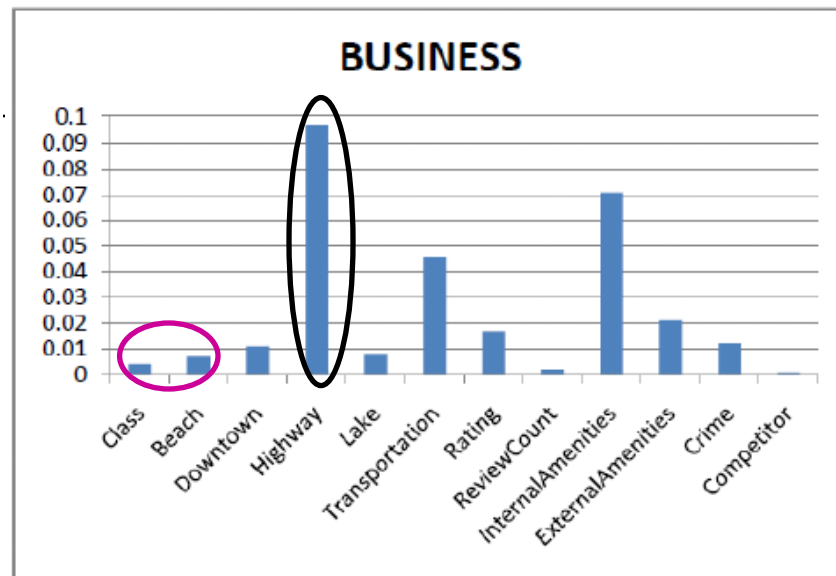
Result (1) - Economic Marginal Effects

Characteristics	Marginal Effect
Public transportation	18.09%
Beach	18.00%
Interstate highway	7.99%
Downtown	4.70%
Hotel class (Star rating)	3.77%
External amenities	0.08%
Internal amenities	0.06%
Annual Crime Rate	- 0.27%
Lake/River	- 12.94%

Result (2) – Preference Deviations Based on Different Travel Purposes



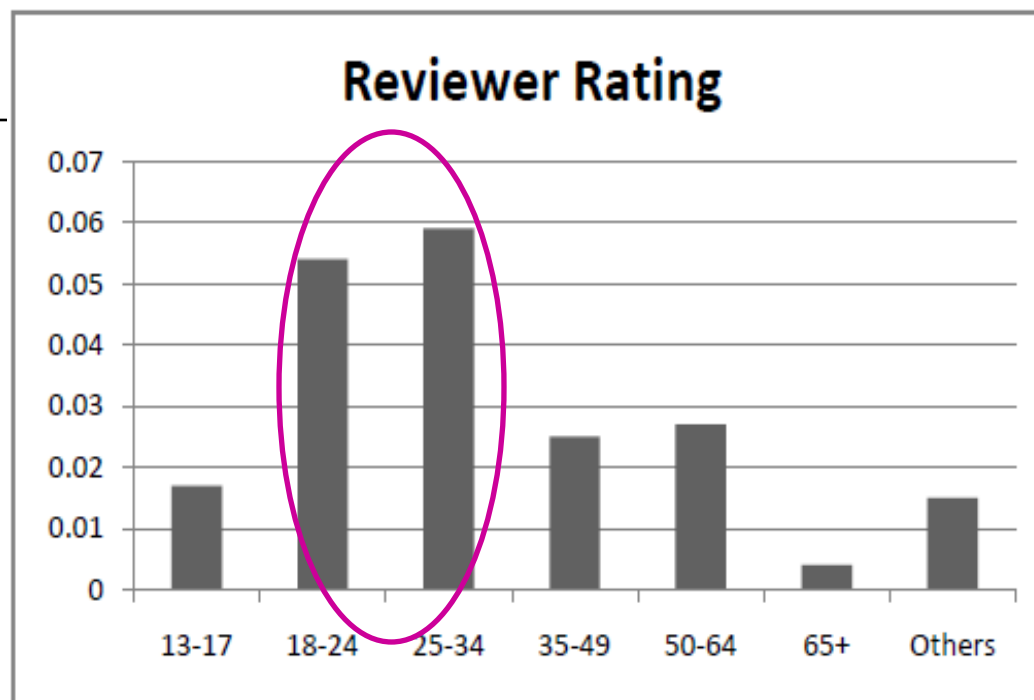
(a)



(b)

Consumers with different travel purposes show different preferences towards the same set of hotel characteristics.

Result (2) - Sensitivity to Online Rating Based on Different Age Groups



Age 18-34 pay more attention to reviews than other age groups.

Ranking Evaluation - User Study (1)

Experiment 1: Blind pair-wise, 200 AMT users, 6 cities, 10 baselines.

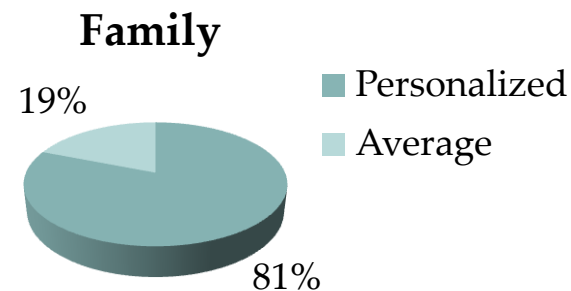
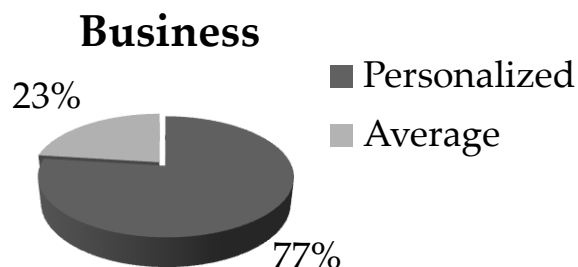
Finding: Our surplus-based ranking is overwhelmingly preferred in any single comparison! ($p=0.05$ sign test)

User Explanations: Diversity; Price not the only factor;
Multi-dimensional preferences.

Our reasoning: Our economic-based model introduces
“diversity” *naturally*.

Ranking Evaluation - User Study (2)

Experiment 2: Blind pair-wise, 200 AMT users.



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%

In all cases, the personalized approach is significantly preferred at the $p = 0.1$ level.

Conclusion & Future Work

A New Ranking System for Product Search

- ✓ Economic utility theory, “Best Value” ranked on Top;
- ✓ Validated with user study with +15000 users, 6 cities.

Major Contributions:

- **Inter-disciplinary** approach
- Captures consumer **decision making process**
- **Privacy-preserving:** Aggregate data → Personal preferences



Future Directions:

- Product bundles
- Consumer browsing data

Q & A

Demo: <http://2.nyuhotels.appspot.com/>

(April 1): 10am-12pm, 1:30-3:30pm

Thank you!

bli@stern.nyu.edu