Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowd-Sourced Content

Beibei Li
(with Anindya Ghose and Panos Ipeirotis)

Stern School of Business
New York University

IDEI – Toulouse, France
Jan-14-2011
I Want to Find…

“How can I find the best hotel in Long Beach?”

Downtown
Beach
Nightlife, Restaurants, Bars…
A Great Price for what it offers!
Limitations of Online Search Engines

Problem:
- Decision of buying a product ≠ judging a document as relevant.
- Largely ignore multidimensional preferences;
- No consumer heterogeneity.
Limitations of Customer Information Seeking

How can we effectively learn consumer preferences from various types of social media data (e.g., UGC and Crowd-Sourced Data) to improve user search experience?
How to improve consumer search experiences with various types of social media data?

**Main Idea:** Identify latent demand pattern by analyzing social media data → Ranking hotels based on consumer surplus → Reduce consumer search costs of evaluating offers online.

**Key Challenge:** Bridge the gap between the qualitative nature of social media content and the quantitative nature of economic choice models.

**Method:** Combine Text Mining, Image Classification, Social Geo-Tagging, Crowd-Sourced User Survey with Structural Modeling for Demand Estimation.

**Validation:** User experiments to validate our ranking system with existing benchmark systems offered by travel search engines.
Agenda

- Data
- Hybrid Structural Demand Model
- Results
- Conclusion
Data

Transaction data:  Travelocity.com, 1497 US hotels, 2008/11-2009/1

Location Characteristics:
- Social geo-tags:  Geonames.org, “Public transportation”
- GeoMapping Search Tools:  Microsoft Virtual Earth SDK, “Restaurants”
- Image Classification:  “Beach”, “Downtown”
- On-Demand Survey:  Amazon Mechanical Turk (AMT), “Highway”

Service Characteristics:
- JavaScript parsing engines:  TripAdvisor & Travelocity, “# of Internal amenities”, “Reviewer Rating”, “# of online reviews”

Additional Review Characteristics:
- Text Mining:  Review-based content from TripAdvisor & Travelocity, Text features (e.g., “Breakfast”, “Staff”), “Subjectivity”, “Readability”, “Disclosure of Reviewer Identity”
Structural Model Framework

- Each hotel exclusively belongs to 1 of the 8 Travel categories:
  - Family Trip, Business Trip, Romantic Trip, Tourist Trip, Trip with Kids, Trip with Seniors, Pet Friendly, and Disability Friendly.

- A consumer makes a choice decision by looking for the hotel with specific service/location characteristics that best match her travel purpose:
  - e.g., A romantic trip traveler prefers:
    1) characteristics: downtown, beach, nightlife, restaurants…
    2) hotels in romance category are best candidates to satisfy her needs.

**Our Model:** Hybrid random coefficient demand model, capturing two levels of consumer heterogeneity during the decision process.
We propose a hybrid random coefficient-based structural model:

\[ u_{ijkt} = X_{jt} \beta_i - \alpha_i P_{jt} + \xi_{jt} + \varepsilon_{ikt}^k, \]

where \( u_{ijkt} \) represents the utility of hotel \( j \) with category type \( k \) (\( 1 \leq k \leq 8 \)) for consumer \( i \) at time \( t \). The \( \varepsilon \) with a superscript \( k \) represents a travel category level “taste shock”.

- Consumer-specific random coefficients:
  \[ \beta_i \sim (\beta, \sigma_\beta), \ \alpha_i \sim (\alpha, \sigma_\alpha) \]

- Choice of travel category
- Choice of hotel within a travel category

**Summary:** Combine the BLP (1995) and PCM (Berry & Pakes 2007).
### Estimation Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef. (Std. Err) I</th>
<th>Coef. (Std. Err) II</th>
<th>Coef. (Std. Err) III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price$^{(I)}$</td>
<td>-0.122*** (.002)</td>
<td>-0.121*** (.003)</td>
<td>-0.119*** (.002)</td>
</tr>
<tr>
<td>CHARACTERS$^{(I)}$</td>
<td>0.009*** (.002)</td>
<td>0.009*** (.002)</td>
<td>0.009*** (.002)</td>
</tr>
<tr>
<td>COMPLEXITY</td>
<td>-0.010*** (.002)</td>
<td>-0.010*** (.003)</td>
<td>-0.010*** (.003)</td>
</tr>
<tr>
<td>SYLLABLES$^{(I)}$</td>
<td>-0.040*** (.006)</td>
<td>-0.041*** (.006)</td>
<td>-0.042*** (.006)</td>
</tr>
<tr>
<td>SMOG</td>
<td>0.076** (.025)</td>
<td>0.075** (.025)</td>
<td>0.077** (.026)</td>
</tr>
<tr>
<td>SPELLERR$^{(I)}$</td>
<td>-0.128*** (.003)</td>
<td>-0.128*** (.004)</td>
<td>-0.128*** (.004)</td>
</tr>
<tr>
<td>SUB</td>
<td>-0.146*** (.007)</td>
<td>-0.149*** (.007)</td>
<td>-0.152*** (.008)</td>
</tr>
<tr>
<td>SUBDEV</td>
<td>-0.420*** (.008)</td>
<td>-0.422*** (.008)</td>
<td>-0.425*** (.009)</td>
</tr>
<tr>
<td>ID</td>
<td>0.054* (.023)</td>
<td>0.055* (.024)</td>
<td>0.056* (.027)</td>
</tr>
<tr>
<td>CLASS</td>
<td>0.032*** (.009)</td>
<td>0.034*** (.009)</td>
<td>0.037*** (.009)</td>
</tr>
<tr>
<td>CRIME$^{(I)}$</td>
<td>-0.023* (.016)</td>
<td>-0.025* (.015)</td>
<td>-0.029* (.015)</td>
</tr>
<tr>
<td>AMENITYCNT$^{(I)}$</td>
<td>0.005* (.016)</td>
<td>0.004* (.002)</td>
<td>0.005* (.002)</td>
</tr>
<tr>
<td>EXTAMENITY$^{(I)}$</td>
<td>0.007*** (.001)</td>
<td>0.008*** (.002)</td>
<td>0.008*** (.002)</td>
</tr>
<tr>
<td>BEACH</td>
<td>0.153*** (.003)</td>
<td>0.156*** (.003)</td>
<td>0.158*** (.003)</td>
</tr>
<tr>
<td>LAKE</td>
<td>-0.110*** (.033)</td>
<td>-0.111*** (.033)</td>
<td>-0.112*** (.033)</td>
</tr>
<tr>
<td>TRANS</td>
<td>0.156*** (.002)</td>
<td>0.159*** (.003)</td>
<td>0.167*** (.005)</td>
</tr>
<tr>
<td>HIGHWAY</td>
<td>0.068* (.028)</td>
<td>0.070** (.025)</td>
<td>0.064* (.027)</td>
</tr>
<tr>
<td>DOWNTOWN</td>
<td>0.040*** (.003)</td>
<td>0.045*** (.003)</td>
<td>0.045*** (.003)</td>
</tr>
<tr>
<td>TA_RATING</td>
<td>0.039** (.017)</td>
<td>0.040** (.018)</td>
<td>0.035* (.018)</td>
</tr>
<tr>
<td>TL_RATING</td>
<td>0.035*** (.009)</td>
<td>0.036*** (.009)</td>
<td>0.035*** (.008)</td>
</tr>
<tr>
<td>TA_REVIEWCNT$^{(I)}$</td>
<td>0.184*** (.045)</td>
<td>0.186*** (.046)</td>
<td>0.182*** (.045)</td>
</tr>
<tr>
<td>TA_REVIEWCNT$^{(I)}$</td>
<td>-0.052*** (.006)</td>
<td>-0.053*** (.006)</td>
<td>-0.052*** (.006)</td>
</tr>
<tr>
<td>TL_REVIEWCNT$^{(I)}$</td>
<td>0.013*** (.002)</td>
<td>0.012*** (.002)</td>
<td>0.012*** (.002)</td>
</tr>
<tr>
<td>TL_REVIEWCNT$^{(I)}$</td>
<td>-0.021*** (.006)</td>
<td>-0.023*** (.005)</td>
<td>-0.024*** (.005)</td>
</tr>
</tbody>
</table>

I. >= 1 review from either TA or TL.
II. Reviews >= 5.
III. Review >= 10.
# Marginal Effects

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Marginal Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public transportation</td>
<td>18.09%</td>
</tr>
<tr>
<td>Beach</td>
<td>18.00%</td>
</tr>
<tr>
<td>Interstate highway</td>
<td>7.99%</td>
</tr>
<tr>
<td>Downtown</td>
<td>4.70%</td>
</tr>
<tr>
<td>Hotel class (Star rating)</td>
<td>3.77%</td>
</tr>
<tr>
<td>External amenities</td>
<td>0.08%</td>
</tr>
<tr>
<td>Internal amenities</td>
<td>0.06%</td>
</tr>
</tbody>
</table>
Robustness Checks

- **Estimation using alternative sample split:**
  Samples consisting of those hotels that have at least one review from only Travelocity or only TripAdvisor or both.

- **Estimation using alternative models:**
  Use BLP model, PCM, Nested Logit, OLS with random effect.

- **Estimation using additional features:**
  “Airport”, “Convention center”, and 7 individual service ratings from TripAdvisor, eg, Value, Room, Location, Cleanliness, Service, Check-in, Business service.

- **Estimation using alternative IVs:**
  Lag Price with Google Trend, Employee wage, Region dummies, BLP style IVs (Average characteristics of the same star-rating hotels in the other markets).

  All the results are qualitatively very consistent with our findings above.
Consumer Surplus Based Ranking

- Ranking hotels based on the **consumer surplus** of each hotel for consumers on an aggregate level.
- How much “**extra value**” consumers can obtain after purchase?

**User Study:** Blind pair-wise comparisons, 200 anonymous AMT users; Compare with 13 existing benchmarks.

**Result:** CS-based ranking is overwhelmingly preferred. More than 80% customers (p=0.001, sign test)

**Reasoning:** - Diversity;
- Price & Quality;
- Multi-dimensional preferences.
<table>
<thead>
<tr>
<th>Cities</th>
<th>Baselines</th>
<th>Rating Tripadvisor</th>
<th>Rating Travelocity</th>
<th>Most Booked</th>
<th>Price Low to High</th>
<th>Price High to Low</th>
<th>Hotel Class</th>
<th># of Reviews</th>
<th># of Rooms</th>
<th># of Amenities</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>77%</td>
<td>63%</td>
<td>61%</td>
<td>57%</td>
<td>71%</td>
<td>88%</td>
<td>76%</td>
<td>89%</td>
<td>60%</td>
<td></td>
</tr>
<tr>
<td>Los Angeles</td>
<td>72%</td>
<td>58%</td>
<td>71%</td>
<td>59%</td>
<td>84%</td>
<td>89%</td>
<td>87%</td>
<td>86%</td>
<td>69%</td>
<td></td>
</tr>
<tr>
<td>San Francisco</td>
<td>79%</td>
<td>57%</td>
<td>65%</td>
<td>62%</td>
<td>70%</td>
<td>82%</td>
<td>68%</td>
<td>79%</td>
<td>79%</td>
<td></td>
</tr>
<tr>
<td>Orlando</td>
<td>83%</td>
<td>81%</td>
<td>62%</td>
<td>63%</td>
<td>73%</td>
<td>79%</td>
<td>73%</td>
<td>79%</td>
<td>61%</td>
<td></td>
</tr>
<tr>
<td>New Orleans</td>
<td>61%</td>
<td>69%</td>
<td>60%</td>
<td>78%</td>
<td>69%</td>
<td>80%</td>
<td>72%</td>
<td>91%</td>
<td>58%</td>
<td></td>
</tr>
<tr>
<td>Salt Lake City</td>
<td>61%</td>
<td>80%</td>
<td>69%</td>
<td>66%</td>
<td>79%</td>
<td>83%</td>
<td>73%</td>
<td>70%</td>
<td>76%</td>
<td></td>
</tr>
</tbody>
</table>

Significance Level: $P=0.05 \geq 62\%$, $P=0.01 \geq 66\%$, $P=0.001 \geq 72\%$ (Sign Test, $N=200$)

Mixed Rating Strategy:
(i) Average of Tripadvisor rating and Travelocity rating when both are available;
(ii) Equal to one of the two ratings if the other one is missing;
(iii) Zero when both ratings are missing.
Can we do better?

- Personalized ranking
  - Incorporate consumer demographics;
    - i.e., age group, travel purpose
  - Examine the interaction effect between consumer demographics and hotel characteristics; for example,
    - Travel Purpose and Price;
    - Travel Purpose and Hotel Characteristics (e.g., location, service, etc.);
    - Travel Purpose and Brands.
  - Derive personalized consumer surplus for ranking.
Weights of Hotel Characteristics Based on Travel Purposes

Consumers with different travel purposes assign different weight distributions on the same set of hotel characteristics.
Selected Findings from Extended Model

• Business travelers have the highest marginal valuation for “highway” and “transportation.”

• Romance travelers have the highest marginal valuation for “hotel class” and “beach.”

• Tourists have stronger preferences towards “Hilton” and “Intercontinental.” Senior citizens have stronger preference towards “Best Western”

• etc.
User Study (2)

Experiment 2: **Blind pair-wise comparisons, 200 anonymous AMT users; baseline: generalized CS-based ranking (for an average consumer).**

*E.g., Business trip and family trip AMT user study results in the NYC experiment.*

Conclusion: Personalized CS-based ranking is overwhelmingly preferred.

**Reasoning:** The personalized ranking model can capture consumers’ specific expectations, dovetail with their actual purchase motivation in the real world.
Conclusions

✓ Economic impact of hotel characteristics using user-generated and crowd-sourced data.

✓ Structural model for demand estimation, image classification, text mining & on-demand social annotations.

✓ Model extension to capture interactions of hotel features with travel purpose and brands.

✓ New ranking system for hotels on travel search engines and validate it with field experiments.

• ➔ Any product search engines

Demo:  http://nyuhotels.appspot.com/
Future Work

In the future, we plan to look into:

• How does the “rank” of hotel affect clicks and conversions for that hotel? What other factors drive “search”, “clicks”, and “conversions” of a hotel?
  - ~1 M online sessions, >40 M search events;
  - Hierarchical Bayesian modeling;

• Why do some people search more, while others search less?
  - Empirically estimate consumer search cost in travel industry;
  - Dynamic structural modeling.