

# Tackling the Long Tail Problem in Online Marketing Optimization

Wenqing Lu

2/19/2014



# Online Marketing for Orbitz



## Google

Chicago, Illinois

Web Flights **Hotels** More

Friday, February 28 Sunday, March 2

Price Hotel class Brand User rating Amenities Location

559 hotels Saved 0 Map

Sort by Relevance

**Hard Rock Hotel Chicago**

4.1 ★★★★★ 264 reviews 4-star hotel

Set in refurbished Chicago landmark building - Rooms are contemporary and feature feather top bedding.

Free WiFi

**\$154** per night

Book

**Intercontinental Chicago Magnificent ...**

4.0 ★★★★★ 229 reviews 4-star hotel

Striking Hotel Complex in Landmark 1929 Building with adjacent high rise Hotel Tower.

Indoor pool

**\$153** per night

Good deal

Book

**The Westin Chicago River**

4.1 ★★★★★ 185 reviews

Elegant Hotel located downtown on River, adjacent to business district

Free breakfast · Free WiFi

Book on **InterContinental.com** \$178 incl taxes & fees

Book on **Orbitz** \$178 incl taxes & fees

See more details

## Trip Advisor

San Diego Hotels 269 of 269

02/18/2014 02/20/2014 Change dates Clear dates

All San Diego Hotels (269) Just for you Best Value (25) Family (11) Business (15) Romantic (3) Green (2) Luxury (2) On the Beach

Price Rating Neighborhood More Sort Availability

**The Grand Del Mar** ★★★★★

Travelers' Choice® 2014 Winner (Luxury) | Top Hotels

Offers & Announcements 3rd Night Complimentary

GreenLeaders Bronze level

#1 of 269 hotels in San Diego

★★★★★ 712 reviews

"Excellent stay in San Diego" 02/15/2014

"great family get away" 02/14/2014

Professional photos | Traveler photos (26) | Map

**Residence Inn San Diego Scripps Poway Parkway**

★★★★☆

Special Offer Earn Rewards Points

#2 of 269 hotels in San Diego

★★★★★ 173 reviews

"Amazing Staff & Beautiful Spacious..." 11/16/2014

"Excellent extended stay" 01/17/2014

Professional photos | Traveler photos (27) | Map

**Expedia**

**\$395\*** per night

amedtravel.com \$499\*

Orbitz.com \$499\*

Travelocity \$399\*

See all 7

**Residence Inn San Diego**

**\$189\*** per night

Orbitz.com \$189\*

Priceline.com \$189\*

OneTravel.com \$189\*

See all 12

Go to map view

Special offers in San Diego

**The Bed and Breakfast Inn at La Jolla**

★★★★★ 141 Reviews

\$159 (USD)

**SpringHill Suites San Diego Rancho**

★★★★★ 136 Reviews

Earn Rewards Points

**Howard Johnson Inn San Diego Hotel**

★★★★★ 211 Reviews

Free Breakfast Tilt

View all Special Offers in San Diego

Top picks in San Diego

**Park Suites**

★★★★★ 21 Reviews

San Diego, CA

Show Prices

# Optimizing Bidding on Travel Sites



- Business objective:
  - Grow transaction volume
  - Be more efficient
- Goal
  - Bid intelligently for each hotel at travel sites
- Key element
  - Estimate the revenue per click (RPC) at each hotel/keyword level

# Predictive Modeling Approach for Bid Optimization



- Predictive modeling approach
  - Building RPC model to calculate the click value at travel sites
  - RPC model enables smart bidding

Illustrative data:

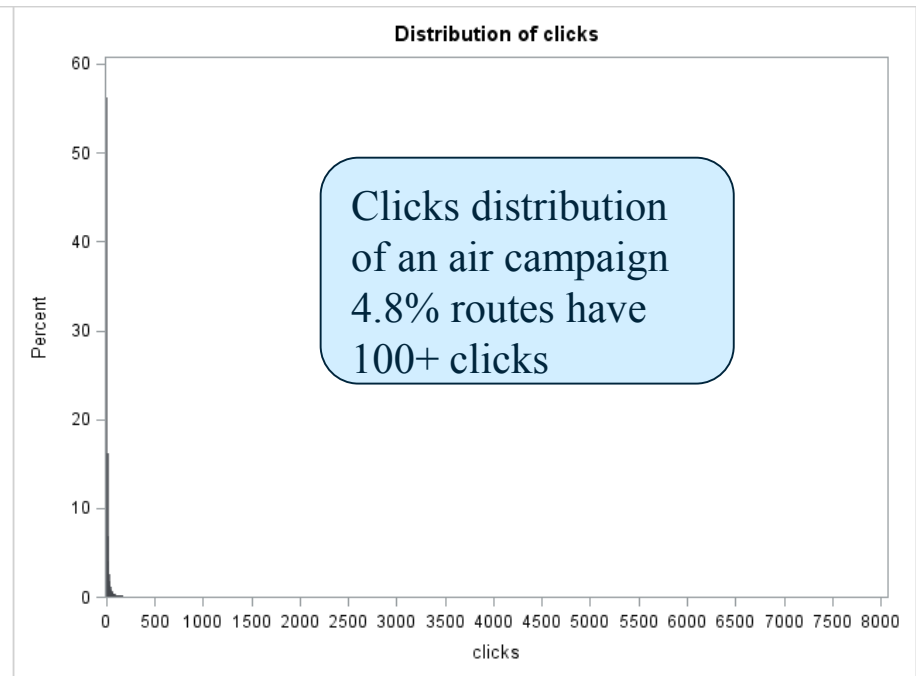
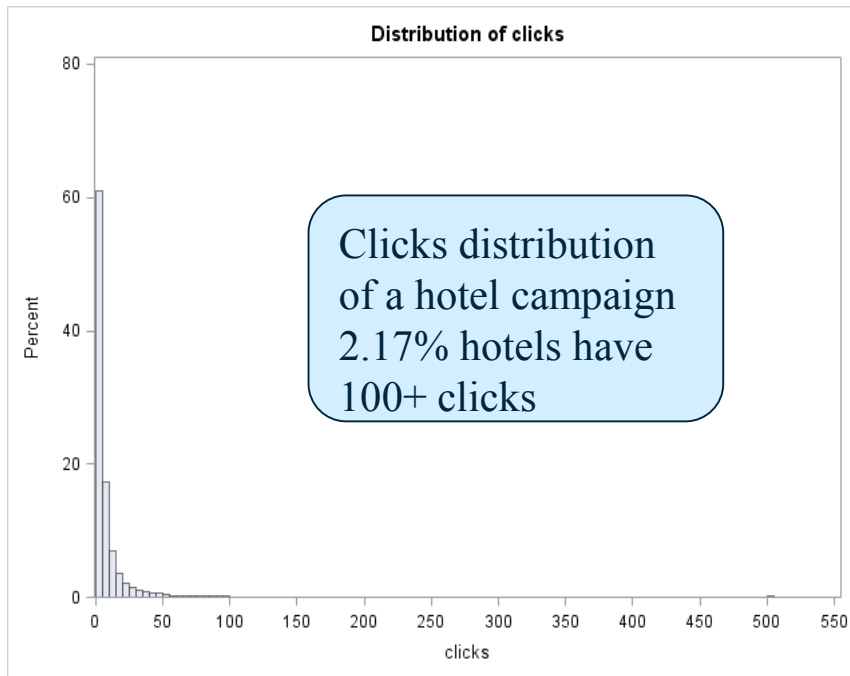
	Flat CPC	Modeled RPC	What to do?
Hotel A	\$ 0.50	\$ 0.20	Reduce CPC
Hotel B	\$ 0.50	\$ 0.75	Increase CPC

# Long Tail Problem and Power Law Distribution



- The power law distribution

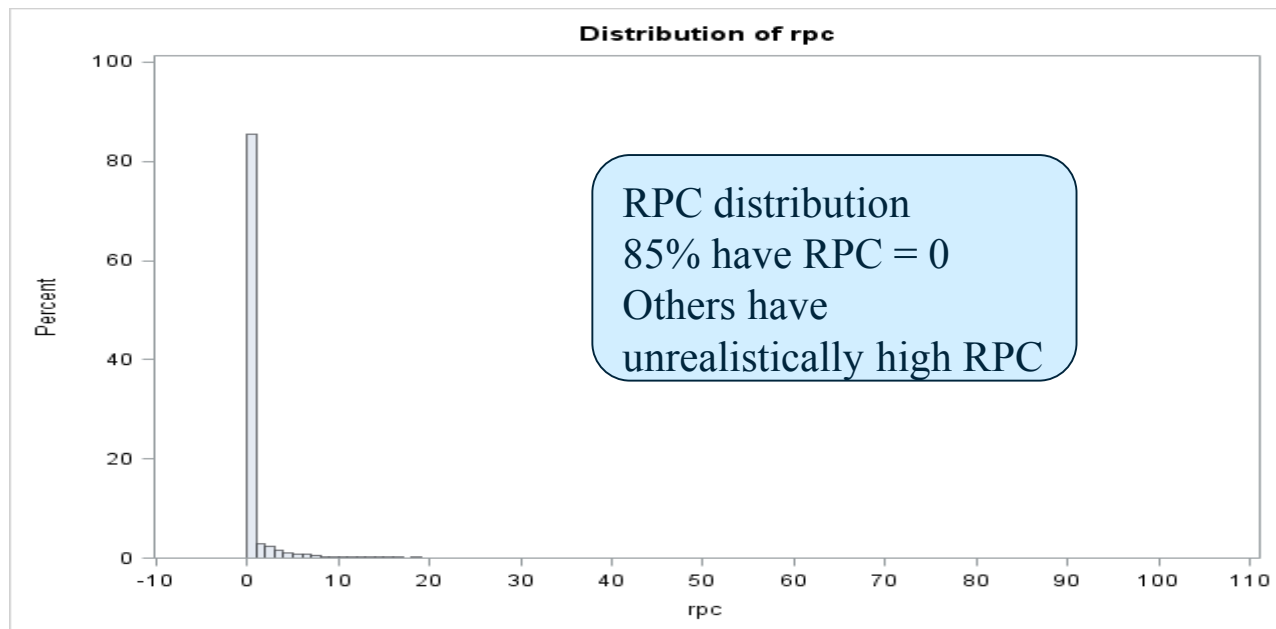
- user search keywords, word frequency, computer file size, and many internet related metrics site traffic



# Sparse Data Problem



- Calculating RPC directly we can only get good estimate for head terms
- RPC for tail will be either
  - Too high for some lucky cases
  - 0 for majority of the other cases



# Challenges Presented by Long Tail Problem



- Tail RPC estimation
  - Model vs. clustering of tail
- Update bidding based on previous performance
  - Each day almost all hotels get only very few clicks
  - How to decide if past day's performance is indicative of future performance
  - How frequent should we update the bid?

# Building a predictive model for RPC



- Building a model for predicting RPC is a hard problem
  - deal with the majority of terms with very sparse data
- Modeling techniques
  - Decision Trees, Regression, Random Forrest, Gradient Boosting Tree, Neural Networks, etc
- Model input variables
  - Hotel properties including location, brand, chain, star-rating, etc.
  - Customer review text can be used as a new source of unstructured data to further improve the prediction power
    - First need to do text mining to turn the unstructured data into numbers



# Bayesian Updating to Combine Model Prediction and Recent Experience



- Suppose we have a Normal Distribution  $N(\mu, \sigma^2)$ , and the mean has a prior distribution of  $N(\mu_0, \sigma_0^2)$ , then the Bayes estimate of the posterior mean is

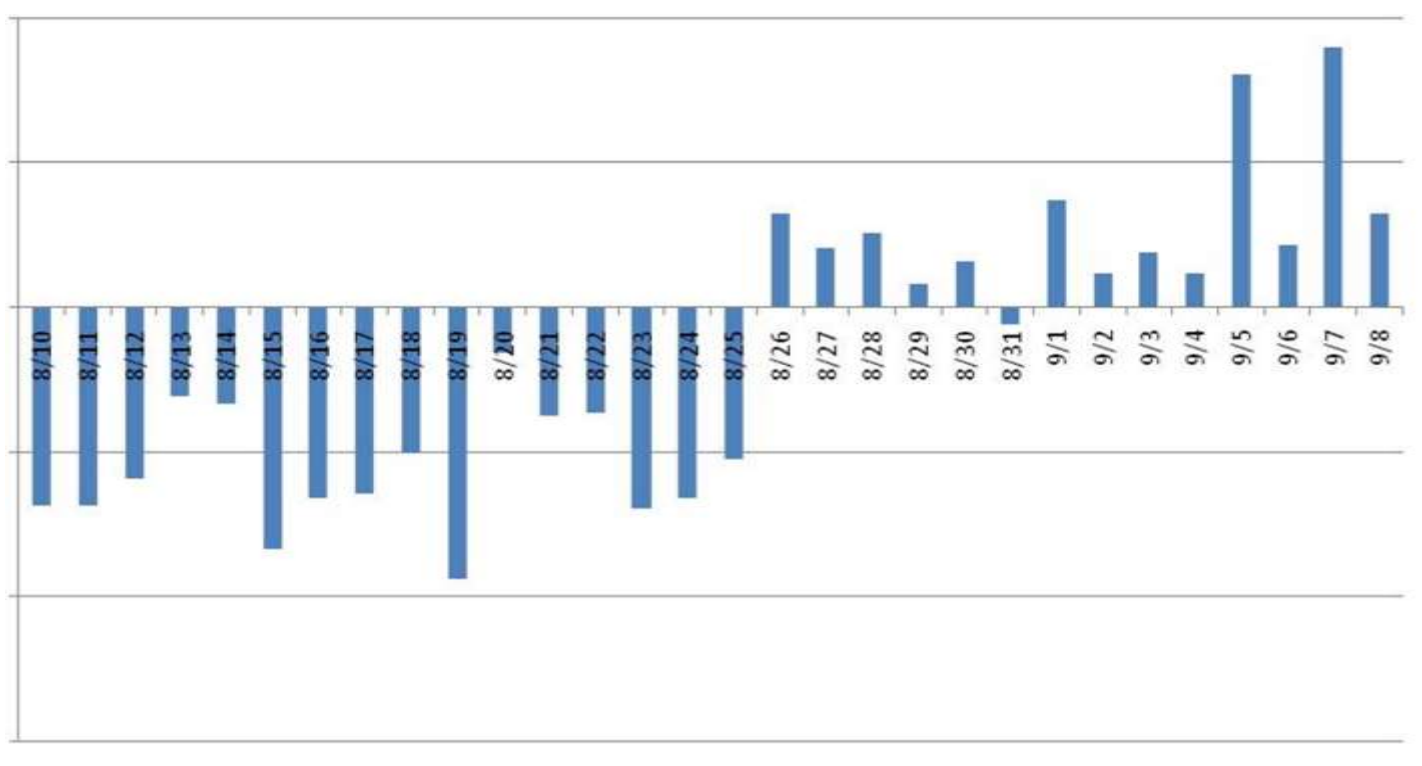
$$\left( \frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n x_i}{\sigma^2} \right) / \left( \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)$$

- Weighted average of prior mean and sample mean
- The weight has three factors
  - The higher confidence for the prior the more weight we give to prior mean
  - The bigger sample size the more weight we give to sample mean
  - The smaller sample variance (estimate of population variance) the more weight we give to sample mean
- We can use this approach to update the RPC model estimates

# Model performance comparison



- Improve the campaign performance significantly
  - From loss to profitability
  - Profitability allows further growth



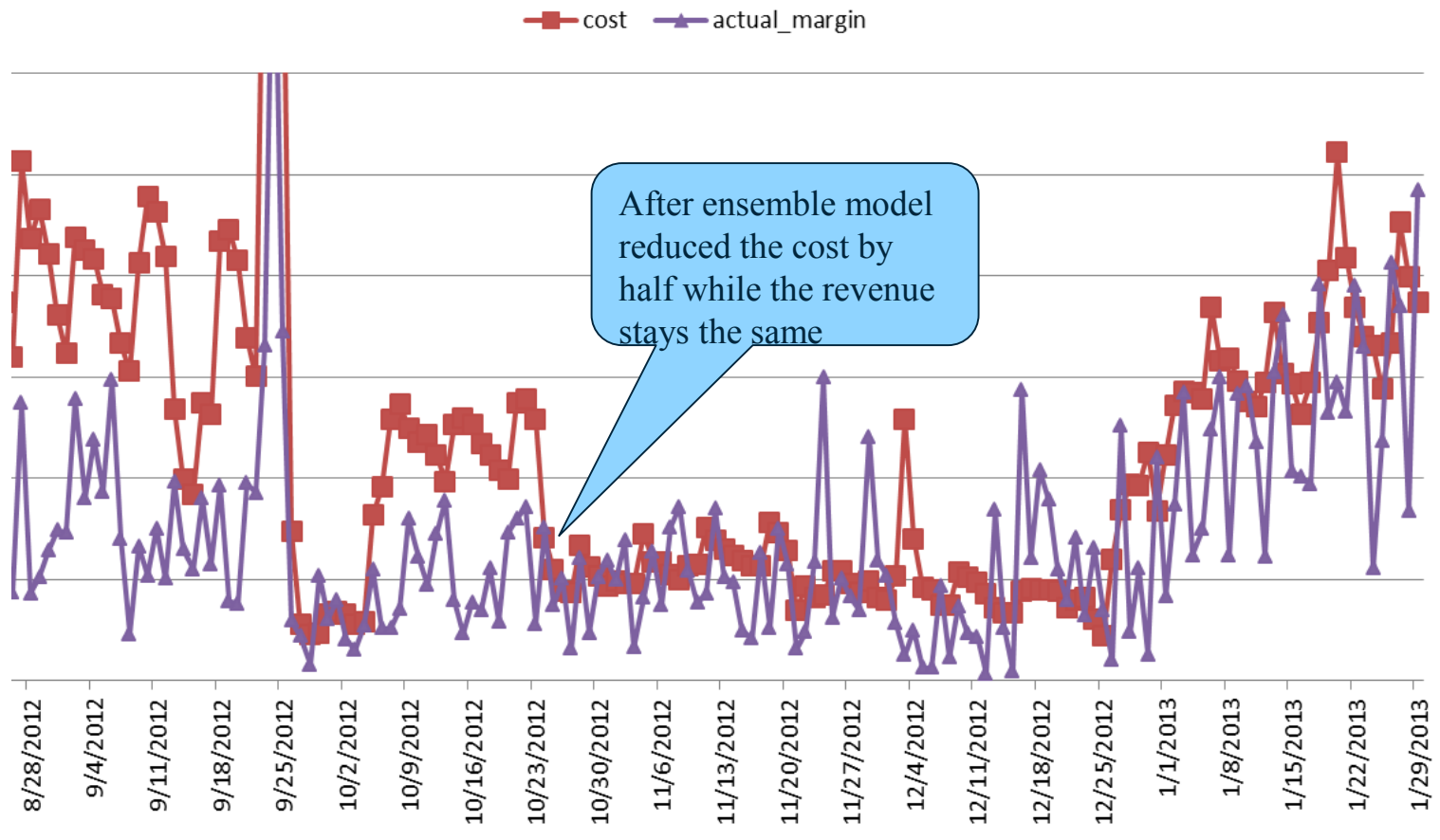


- Ensemble is another technique that can help dealing with sparse data
  - Ensemble model can help with big data and sparse data
  - Bagging and boosting
  - Mixture of different models
- The key for ensemble model is diversity of individual models
- Another factor is how to combine the individual models

# Ensemble model improvement



## UK TA Weekly Performance 2012



# Thanks!



- Email: [wenqing.lu@orbitz.com](mailto:wenqing.lu@orbitz.com)
- LinkedIn [www.linkedin.com/in/wenqinglu/](http://www.linkedin.com/in/wenqinglu/)
- Twitter [@wqlu2010](https://twitter.com/wqlu2010)
- Blog <http://wqlu2010.wordpress.com>