Leveraging Data to Power Local Commerce

Rajesh Parekh
Data Science Team
Groupon: Introduction
From humble beginnings …

Pioneered the daily deals business in October 2008
... to a local commerce leader

More than 38MM active customers globally...

In 48 countries and 500+ markets...

With more than 250,000 merchants served...

1,000+ deals per day across multiple channels – local, goods, travel, …
Satisfaction Benchmarks
ForeSee tracks Satisfaction for over 500 companies

- Groupon’s Merchant Satisfaction score is very strong – especially for a B2B score
- Groupon’s Customer Satisfaction also very strong – Top 5 of Internet retailers*

*Based on ForeSee Satisfaction Study of Top Online Properties (June 2012)
Key Data Science Problems
Identify Great Merchants

Merchants that delight users
Structure the Right Deal

Determine optimal deal structure
Right Slate of Deals

- Restaurants (X%)
- Health and Beauty (Y%)
- Activities (Z%)

Optimize the deal mix for each market
Relevance & Targeting

Target the right deal to the right user(s)
Groupon’s Approach

- Leverage data to drive business decisions
- Rapid: Idea – Experiment – Analyze – Deploy cycle
Optimizing Deal Category Mix
Deal Category Mix

• Problem:
  – Determine the right deal category mix for each market

• Importance:
  – Closing deals with great merchants takes time and effort
  – Sales bandwidth is limited
  – Naïve solutions are suboptimal

• Approach:
  – Deal category mix modeled as an optimal asset allocation problem
  – Find the deal category mix that
    ➤ Achieves a target performance while minimizing risk
    ➤ Preserves category diversity
Portfolio of Assets

• Given a set of assets $A_1, A_2, \ldots A_N$ with:

  \begin{align*}
  \text{Expected Return: } &E(r_i) \quad i = 1,2,3,\ldots,N \\
  \text{Covariance: } &\sigma_{i,j} \quad i,j = 1,2,3,\ldots,N
  \end{align*}

• Define portfolio as weighted allocation of above assets

  \begin{align*}
  \text{Weights: } &w = \{w_1, w_2, w_3, \ldots, w_N\} \\
  \text{Portfolio Return: } &E(r) = \sum_{i=1}^{N} w_i \cdot E(r_i) \\
  \text{Portfolio Risk: } &\sigma^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{i,j}
  \end{align*}

  Markowitz Portfolio Model
Obtaining the Optimal Asset Allocation

Objective:
\[
\min_w \sigma^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij}
\]

Constraints:
\[
s.t. \sum_{i=1}^{N} w_i * E(r_i) = E(r)
\]
\[
\sum_{i=1}^{N} w_i = 1
\]
\[
0 \leq w_i \leq c_i \text{ where } c_i \leq 1
\]

Solving the QP gives the minimum variance frontier
Optimizing the Deal Category Mix

- **Key Idea:**
  - Modeled as a portfolio of *assets* where each deal category is an asset
  - Return and risk determined using historical deal performance
  - Find the allocation \( w \)
  - \( w_1 \)% Restaurant + \( w_2 \)% Health & Beauty + \( w_3 \)% Activities + …

\[
\sum_{i=1}^{N} w_i = 1
\]

- Constraints on individual \( w_i \) are selected to
  - preserve deal diversity
  - reflect market conditions / seasonality
Efficient Frontier

\[ E(r) \]

Low Risk Allocation

Efficient Frontier

Example Historical Allocation

Higher return allocation available for same risk
Deal Relevance and Targeting
Deal Relevance and Targeting

• Show the right deal to each user
• Right experience on web, mobile, email

But wait! Isn’t this already solved in Computational Advertising?
Computational Advertising

- **Search Advertising**
  - Relevant to the search context
  - Slate of ads is very similar (barring spam)
  - Optimize for clicks
  - Long running ads
  - Typically a tight price range
    - (Max – Min) bid is small

- **Contextual Advertising**
  - Relevant to the page context
  - Display ads
  - Optimize for clicks (or brand value)
  - Long running ads
  - Pricing based on content popularity
Nuances in Computing Deal Relevance

• **User is the main context**
  – No other context like *search keywords* or *content* usually available

• **Diverse slate of deals**
  – Don’t feature multiple *pizza* deals in a local market on the same day

• **Wide price ranges**
  – Restaurant deals for $10+ versus Lasik deals for $1000+

• **Short duration**
  – Interesting challenge to optimize deal performance mid-flight

• **Optimize for purchases / conversions**
  – Clicks are only proxies of user interest, brand value is more ephemeral
Predicting Deal Relevance

Features

- User
- Deal
- Merchant

Optimization Function

Algorithms

- Bayesian
- Logistic
- ...

Relevance

Simpler algorithms trained with large amount of data
Optimization Function

Non-trivial Choice

Conversion

\[ P(\text{conversion}) \]
- Favors lower price deals

Revenue

\[ E(\text{rev}) = P(\text{conversion}) \times \text{price} \]
- More expensive deals can dominate

Need to balance multiple, often conflicting objectives
Modeling Insights: Location

Distance to deal location is key
Modeling Insights – Real-Time Performance

- Deal Performance picks up during the day

- Early performance strongly correlated with overall performance

Leverage early deal performance to optimize targeting
Conclusion

• **Rich technical challenges**
  – Systems and scalability
  – Predictive Modeling, Data Mining, and Machine Learning
  – Combinatorial optimization

• **Unique interplay of web, mobile, email, and social**
  – Potential to make significant impact to the business

• **Small, tight knit, high energy team**
  – Degrees in CS, Applied Math, Stats, Physics, …
We’re Hiring! – [http://www.jobs.groupon.com](http://www.jobs.groupon.com)

- web application development
- mobile design & development
- social engagement
- client-side tech, design, UX
- high-performance distributed systems
- algorithms & optimization
- systems engineering & operations
- development tools & quality innovation
- data science & analytics
- databases & data warehousing
- business strategy innovations
- ... and more ...

CIKM 2012
Acknowledgements

Thanks to many talented individuals at Groupon I am privileged to work with!

- Data Science
- Engineering
- Marketing / Market Research
Questions?

Rajesh Parekh
Groupon
Director, Research
rajesh@groupon.com