Frontiers in E-Commerce Personalization

Sri Subramaniam
VP, Relevance, Groupon
(srisub@groupon.com)
E-Commerce 1.0 → 2.0: Still “Pull”

- **E-commerce 2.0** seems to be about: insanely bigger catalogs, insanely fast shipping and user experience
- **Search** is the dominant finding paradigm
- Weak-intent screens (ex: Home, Browse) have been polluted by Ads, leading to a self-fulfilling tune-out with consumers
- Average **basket size** remains under 1.5

- What happened to the **fun** experience of walking down an aisle in a mall store, and picking up stuff you didn’t think about?
- Something’s missing…
E-commerce 1.0 → 2.0: Discovery

• Make online shopping more **delightful**, more like a game
• Bring back **curation** and careful selection of inventory
• **Deliberate serendipity**: make consumers feel they stumbled upon something cool
• Under the hood, massive use of science, and context

• **Context = Mobile, Social**
  • Your phone just knows so much about you – locations, likes, pins, apps, products you browse
  • Fire up “fun shopping app”: voila, the 5 deals you are most likely to buy on impulse, at that time, at that place!
Local: Ripe for Discovery

• Relatively **low inventory** (100Ks, not 100Ms)
• Consumers have **persistent interests**, but are open to serendipity and adjacency
• Consumers are open to new deals around same interests

• Mobile + Local:
  • Consumers are hooked to their devices, creating a persistent connection to exploit
  • We can learn so much more about the **user**
  • **Impulse buying** can be huge, if the consumer can be inspired!
Leading the way in mobile commerce

Groupon’s vibrant mobile marketplace connects consumers with their local economy

Nearly **92 million** people worldwide have downloaded our mobile app at the end of Q2 2014, including more than 33 million in 2013.

More than **half** of our global transactions were completed on a mobile device by the end of Q2 2014

*Our mobile app is available in 43 countries.*

One of the **25** most downloaded free apps of all time

Right deals to the right users using the right medium at the right time
Objective Function

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
Deal Performance

- Merchants are expecting to be heavily “featured” on the first week of running a new deal
- Mobile: glean deal performance within an hour of launch
- Harder in e-mail: need 2-day “pre-feature” period

- **Explore / Exploit**: give deals a chance, with no pre-disposition
- **Signals**: Views, Add-to-Carts, Purchases over impressions at specific positions
- Measure **correlation with user attributes** – Male/Female, Age Group, Engagement stage, etc
- **Time decay** applied to emphasize recent signals
Local: Location, Location, Location

• Why not sort by distance? Nope!
• Its about “propensity to travel”: how far is a user in a location willing to travel to a deal?

• Varies by:
  • **Category**: 10 miles for Pizza, 100 miles for LASIK
  • **Pop density**: 2 miles for pizza in downtown NY vs 20 miles for pizza in Fargo, ND
  • **Socio-economic**: Less affluent → Affluent, but less of the other way
  • **Suburban/Urban**: Brooklyn → Manhattan, not opp.
  • **Travel / Vacation**: Distance from downtown / hot-spots, not user’s hometown
  • **Lat-long**: Home address vs current GPS location
Local: Location, location, location!

Distance to deal location is key
Location: Zip Affinity scoring

- **Zip affinity** for a deal: Compute affinity with the zip codes from which users are willing to travel to it
  - = **Evidence**: User activity with deal (views, purchases): compute heat map, cut off long tail, normalize
  - + **Priors**: user activity with deals of same category & price range in that zip. Ex: Nail salons in 95129 in price range $30-$50

- **Reverse index**: Given a user’s location, filter list of deals that the user may have a propensity to travel to
- **Distance factor**: Use zip affinity to weight the scoring

- Zip codes capture geo-, socio-economic factors.
  - **Graticules** are less attractive alternative
Location: advanced ideas

• Mobile: glean “hangout” locations:
  • Sample GPS location at night time: probably home
  • Sample at day time: probably work

• Hot spots (restaurants, bars, etc): often not in zip code, or near home. Need to consider nearest (hip) “downtown” location

• Commute from home to work?

• Notify when user is near a deal
  • UX is key to annoying vs delightful
User – Deal – Context Matching

- **Content-based techniques** work well for Local:
  - Persistent interests, transient deals
  - Allows for mixing in user interests from outside (ex: Facebook Likes, etc.)
  - Content-based reco combined with deal performance

- User attributes (**UA**): Gender, Age group, engagement stage, etc
- Deal attribute (**DA**): categories, semantic tags, price range, etc
- Context attributes (**CA**): Time of day, Season, Occasion, etc

- **User activity** (Search, Browse, View, Purchase, etc)
- **Cold-start** problem mitigated by:
  - Explicit personalization, Facebook profile mining
  - Priors: UA-DA tables
Deal Features

• **Semantic tags:**
  - Category hierarchies (Ex: Fitness → Gyms)
  - Descriptive Tags (Ex: Romantic, Good for kids, Gift, etc)
  - Entity tags (Ex: Panini, Scalp massage)
  - Attributes (Ex: Hotel wi-fi, etc)

• Tags can be learnt: how gift-worthy is this item?
• Price range, Deal recency, etc

• **User activity** (View, Purchase, etc) with deals, adds score for attributes DA
• **Browse & Search** (query-categorized) into DA scores
• **Averaging** over all users: allows for unique interests to shine
• Learn weights for various attribute groups, activities
Semantic graphs

• Unified graphs can be useful to represent all kinds of information:
  • Nodes and relationships: Is-a, Contains, Also-Purchased-with, etc

• Hierarchy allows progressively refined personalization

• Allows “deliberate serendipity”

![Concept Graph]

- Restaurants
- Sushi
- Pizza
- Wine Tasting
- Chicken Wings (tag)
- Foodie
- child
- also bought
Collaborative Filtering

• Ephemerality of deals forces us to look at deal attributes
• CF can be applied to deal attributes
  • People who like Pizza also like Italian
• Use purchase/view activity of users to find affinity of deal attributes to each other
• Item-Item CF can be layered on top, esp. for popular deals

• DA-DA CF allows us to extend the user’s feature vector – helps with “semi-active” users
User Features

- Gender (Male vs Female) is a stronger signal for Local, compared to e-commerce goods
  - Age group, income level, etc: not very strong
- Engagement stage (New user, active, inactive, “looker”)
  - New users tend to buy Restaurant deals
  - Active users tend to look for latest deals

- Priors: Compute table of UA-DA correlation
  - Ex: (Women, Jewelry) vs (Women & Men, Hotels)
- Gender can be tricky
  - Females tend to buy for household, males tend to look at, but not purchase female deals
User Profiling

• *The How*: Thin line between creepiness and delight
  – Transparency and explanations allow you to push the frontier

• Cold start problem:
  – Facebook profile: Demographic, glean interests from Likes, even posts
  – Behavioral targeting, but has privacy concerns
  – Location, gender and age give you a headstart

• Explicit personalization: “What do you like?”
  – Choosing from a list of interests has poor adoption
  – Doesn’t work well. “Healthy living” likers end up buying pizza!
  – Explicit dislike works better. “What do you hate?”, “Don’t show me deals like this.. Ever”
Diversity

• Matching brings focus, so need to diversify results to mix it up a bit
• Important for discovery: homogeneity causes drop-off in user interest!
• Multiple dimensions: product mix, categories, price range, etc
• Should be done along adjacent sliding windows of deal results

• Note that any diversity will reduce "pure" relevance
Freshness and “Back-off”

- Active users need to see fresh, new deals every time
- Lesser the intent, more the need for freshness
- But.. If viewed or added to cart, show it MORE – retargeting

- Freshness: Back-off or de-boost based on last set of impressions
- Back-off from entire categories if user is not showing interest

- Purchase Backoff
  - If user purchased something, back-off for a period that depends on category of item
    - 2 weeks for pizza
    - 100 years for Lasik surgery!
Personalized Search

• In a low-intent session, “Search” is not that targeted either
  • Ex: “male grooming”, “spa package”, “things to do”

• Personalization can boost relevance significantly
  • “things to do”: show more Kids Activities to a mom
  • “fragrance”: show men’s perfumes to a man

• “Banded” broad-matching: addresses openness of user to explore, even in Search:
  • Search results can be shown in bands of decreasing relevance
  • Personalized ranking within each band
  • Each band represents a broader match:
    • Ex: “Sky Zone” → “Sky Zone”, followed by “trampoline, bounce house”, followed by “Kids activities”
Summary

- E-commerce “2.0” should be as much about personalization, as it's about speed and comprehensiveness.
- Local commerce (such as Groupon) relies heavily on personalized recommendations.
- Location-awareness is a huge component of Local.

- Content-based techniques work well for Local, and persistent interests in general.
  - But to be truly effective, content-based techniques must be combined with classic performance signals and CF.
  - Diversity and freshness become important when user intent is low, and they are looking to be delighted.
Acknowledgements

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

- Personalization / Relevance Engineering
- Data Science
- Marketing / Market Research

srisub@groupon.com