Recent Advances and Challenges in E-Commerce
Search & Recommendation Systems

Liangjie Hong
February 3, 2020
Agenda

1. E-Commerce Search and Recommendation
2. Challenge I: Recommendation and Search Eco-System
3. Challenge II: User Intent Understanding
4. Summary
How E-Commerce Search and Recommendation Differ from Their Classic Counterparts?
Observation I: Purchase Decisions are Impacted by Many Factors and are Results of Complex Processes.

- Multiple Pages and Modules
- Multiple Sessions
- Multiple Devices
Observation II: Many Users are Passive and Spontaneous.

- Massive Amount of Noise Interaction Data
- Impulse Purchases
- Non Repeating Behaviors
Recent Publications

**Evaluation and Experimentation**

**Search and Recommendation**

**Machine Learning Systems**
Challenge I: Recommendation and Search Eco-System
Example I:
An A/B Test Result for A New Recommendation Algorithm

<table>
<thead>
<tr>
<th></th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommendation Clicks</td>
<td>+5%</td>
</tr>
<tr>
<td>Search Clicks</td>
<td>-3%</td>
</tr>
<tr>
<td>Revenue</td>
<td>~</td>
</tr>
</tbody>
</table>
Example I: An A/B Test Result for A New Recommendation Algorithm

<table>
<thead>
<tr>
<th></th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommendation Clicks</td>
<td>+5%</td>
</tr>
<tr>
<td>Search Clicks</td>
<td>-3%</td>
</tr>
<tr>
<td>Revenue</td>
<td>~</td>
</tr>
</tbody>
</table>

**Notes:**
1. Improvements might come as a result of a series of A/B testing results.
2. Not shipping early corner-stone results might lead to a sub-optimal user experience in a long run.
3. Shipping placebo results might lead to a sub-optimal user experience in a long run.
Example II: An A/B Test Result for A New Recommendation Algorithm

<table>
<thead>
<tr>
<th></th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommendation Clicks</td>
<td>-10%</td>
</tr>
<tr>
<td>Search Clicks</td>
<td>+5%</td>
</tr>
<tr>
<td>Revenue</td>
<td>+1%</td>
</tr>
</tbody>
</table>
Example II: An A/B Test Result for A New Recommendation Algorithm

<table>
<thead>
<tr>
<th></th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommendation Clicks</td>
<td>-10%</td>
</tr>
<tr>
<td>Search Clicks</td>
<td>+5%</td>
</tr>
<tr>
<td>Revenue</td>
<td>+1%</td>
</tr>
</tbody>
</table>

Notes:
1. Deteriorations might come as a result of a series of A/B testing results.
2. Once damage is done, it might impact machine learning algorithms in many ways (e.g., training bias).
3. Not shipping early corner-stone results might lead to a sub-optimal user experience in a long run.
4. Shipping placebo results might lead to a sub-optimal user experience in a long run.
We need to understand the **interplay** between recommendation and search modules as well as their whole **ecosystem** to create a **coherent** user experience and optimize user engagement.

- **Opportunity 1:** Understand experimental results while multiple teams work on different recommendation and search modules.
We need to understand the **interplay** between recommendation and search modules as well as their whole **ecosystem** to create a **coherent** user experience and optimize user engagement.

- **Opportunity 1:** Understand experimental results while multiple teams work on different recommendation and search modules.
- **Opportunity 2:** Develop and implement strategies to improve multiple modules and possibly optimize overall user engagement.
We need to understand the interplay between recommendation and search modules as well as their whole ecosystem to create a coherent user experience and optimize user engagement.

- **Opportunity 1:** Understand experimental results while multiple teams work on different recommendation and search modules.
- **Opportunity 2:** Develop and implement strategies to improve multiple modules and possibly optimize overall user engagement.
- **Opportunity 3:** Develop machine learning models to directly optimize user engagement from a whole user journey perspective.
We need to understand the **interplay** between recommendation and search modules as well as their whole **ecosystem** to create a **coherent** user experience and optimize user engagement.

- **Opportunity 1:** Understand experimental results while multiple teams work on different recommendation and search modules.
- **Opportunity 2:** Develop and implement strategies to improve multiple modules and possibly optimize overall user engagement.
- **Opportunity 3:** Develop machine learning models to directly optimize user engagement from a whole user journey perspective.

A product change could *induce* changes in user interaction with other products.

- An improved recommendation module could effectively suggest items that satisfy users’ needs so that users don’t need to search as much as usual.
- The overall performance of an improved recommendation module could be *cannibalized* by the induced reduction of user engagement in search.
- The performance of search could be cannibalized by an improved recommendation module.
Common Solution
Splitting Average Treatment Effect (ATE) into Two Parts: Direct Effect and Indirect Effect
Common Solution
Splitting Average Treatment Effect (ATE) into Two Parts: Direct Effect and Indirect Effect

Notes:
1. Causal Mediation Analysis (CMA) is a formal statistical framework to conduct such analysis.
2. Average Direct Effect (ADE) is the direct impact of new recommendations while keeping search behavior fixed.
3. Average Causal Mediation Effect (ACME) is the impact of induced changes in search behavior due to changes in recommendation algorithm.
Common Solution
Splitting Average Treatment Effect (ATE) into Two Parts: Direct Effect and Indirect Effect

Notes:
1. ATE, ADE and ACME has been studied extensively in the literature.
2. Existing methodologies cannot be easily utilized due to violations of the key assumptions in the literature: *no unmeasured causally-dependent mediator*.
3. A typical E-commerce site could have hundreds of web-pages and modules, and all of them could be mediators. It is difficult to measure all of them.
4. We extended ADE and ACME to Generalized ADE (GADE) and Generalized ACME (GACME) respectively.
5. It is easy to implement and only requires solving two linear regression equations simultaneously.
# Case I: RecSys Listing Page Same-Shop Experiment

<table>
<thead>
<tr>
<th>Effect</th>
<th>% Change</th>
<th>Conversion Rate</th>
<th>GMV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GADE</strong> Direct Effect of the Change of Rec Module</td>
<td></td>
<td>0.4959%*</td>
<td>0.1681%</td>
</tr>
<tr>
<td><strong>GACME</strong> The Effect of the Induced Change of Search</td>
<td></td>
<td>-0.2757%***</td>
<td>-0.4200%***</td>
</tr>
<tr>
<td><strong>ATE</strong></td>
<td></td>
<td>0.2202%</td>
<td>-0.2518%</td>
</tr>
</tbody>
</table>

**Notes:**
1. % Change = Effect/Mean of Control
2. ‘***’ p<0.001, ‘**’ p<0.01, ‘*’ p<0.05, ‘.’ p<0.1. Two-tailed p-value is derived from z-test for H<sub>0</sub>: the effect is zero, which is based on asymptotic normality.
Case II:
RecSys Listing Page Internal – Bottom Desktop Experiment

<table>
<thead>
<tr>
<th>Effect</th>
<th>Conversion Rate</th>
<th>GMV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GADE</strong> Direct Effect of the Change of Rec Module</td>
<td>0.3448%*</td>
<td>0.0659%</td>
</tr>
<tr>
<td><strong>GACME</strong> The Effect of the Induced Change of Search</td>
<td>-0.0570%</td>
<td>-0.0926%</td>
</tr>
<tr>
<td><strong>ATE</strong></td>
<td>0.2878%</td>
<td>-0.0267%</td>
</tr>
</tbody>
</table>

**Notes:**
1. % Change = Effect/Mean of Control
2. ‘***’ p<0.001, ‘**’ p<0.01, ‘*’ p<0.05, ‘.’ p<0.1. Two-tailed p-value is derived from z-test for H₀: the effect is zero, which is based on asymptotic normality.
Takeaways

Learnings

• Understanding direct vs. indirect effects enables us to understand the competition between recommendation modules and search results; and give more informed decisions during roll-outs.
• Develop better recommendation strategies such as suggesting items and categories not searched organically or diverse information shown in different surfaces.
• Develop better offline evaluation framework to incorporate both search and recommendation results.
Challenge II: User Intent Understanding
In Amazon, users’ shopping preferences are dynamic and can reflect reoccurring occasions (festivals, holidays, seasonal activities). We can detect occasion-based shopping trends from crowd behavior.
Etsy Users’ Impulse Purchase

At Etsy, users purchase items that are not related to their previous behaviors due to many reasons.
We need to understand an individual’s shopping needs including short-term, long-term, periodical, impulse and inspirational intents to optimize user engagement.

- Opportunity 1: Understand and develop models to tackle the change of an individual’s shopping intent due to external events or occasions that deviate from her long-term interests.
We need to understand an individual’s shopping needs including short-term, long-term, periodical, impulse and inspirational intents to optimize user engagement.

- **Opportunity 1:** Understand and develop models to tackle the change of an individual’s shopping intent due to external events or occasions that deviate from her long-term interests.
- **Opportunity 2:** Understand and develop models to tackle the change of an individual’s shopping intent due to life events (e.g., new babe, house move, graduation and etc.)
We need to understand an individual’s shopping needs including short-term, long-term, periodical, impulse and inspirational intents to optimize user engagement.

- **Opportunity 1:** Understand and develop models to tackle the change of an individual’s shopping intent due to external events or occasions that deviate from her long-term interests.
- **Opportunity 2:** Understand and develop models to tackle the change of an individual’s shopping intent due to life events (e.g., new babe, house move, graduation and etc.)
- **Opportunity 3:** Understand and develop models to inspire an individual’s shopping intent and encourage a user to conduct impulse purchases.
We need to understand an individual’s shopping needs including short-term, long-term, periodical, impulse and inspirational intents to optimize user engagement.

- **Opportunity 1:** Understand and develop models to tackle the change of an individual’s shopping intent due to external events or occasions that deviate from her long-term interests.
- **Opportunity 2:** Understand and develop models to tackle the change of an individual’s shopping intent due to life events (e.g., new babe, house move, graduation and etc.).
- **Opportunity 3:** Understand and develop models to inspire an individual’s shopping intent and encourage a user to conduct impulse purchases.

Etsy Users’ Shopping Preferences

Recommending temporally popular items works better than recommending general popular items when there is an intense shopping trend for a specific occasion.
Etsy Users’ Shopping Preferences

Time gap between purchases for Wedding and Anniversary within a year. More than 50% of purchases for anniversary are near the date of wedding purchase within a time window less than 30 days.
Etsy Users’ Shopping Preferences

The reasons an infant’s items shopper changes his/her shopping behaviors.
Modeling Global and Personal Occasions

\[ o = \sum_{l=1}^{L} \alpha_{ql} v_l, \quad \text{where} \quad \alpha_{ql} = \frac{\exp(s(q, k_l))}{\sum_{l=1}^{L} \exp(s(q, k_l))} \]

Attention Mechanism
Modeling Global and Personal Occasions

Query: $\hat{m}_{p_d}^Q$

Scoring: $s(q, k_j) = \frac{qk_j^T}{\sqrt{D}}$

(Key, Value): $(\hat{m}_{p_1}^K, \hat{m}_{p_1}^V), (\hat{m}_{p_2}^K, \hat{m}_{p_2}^V), \ldots, (\hat{m}_{p_d}^K, \hat{m}_{p_d}^V)$
Modeling Global and Personal Occasions
Modeling Global and Personal Occasions

\[ \text{Query: } \hat{t}^{Q''}_{t_{d+1}} \quad (\text{Key, Value}) : (\hat{t}_1, r_1), (\hat{t}_2, r_2), \ldots, (\hat{t}_M, r_M) \]
Modeling Global and Personal Occasions

The proposed Occasion-Aware Recommendation (OAR) model
Modeling Global and Personal Occasions

Comparison of Different Models. * indicates that the improvement of the best result is statistically significant compared with second best result with $p < 0.01$.

<table>
<thead>
<tr>
<th>Model</th>
<th></th>
<th>Etsy</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Amazon</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NDCG</td>
<td></td>
<td>HR</td>
<td></td>
<td>MRR</td>
<td>NDCG</td>
<td></td>
<td>HR</td>
<td></td>
<td>MRR</td>
<td>NDCG</td>
<td></td>
<td>HR</td>
<td></td>
<td>MRR</td>
</tr>
<tr>
<td></td>
<td>K=5</td>
<td>K=10</td>
<td>K=5</td>
<td>K=10</td>
<td></td>
<td>K=5</td>
<td>K=10</td>
<td>K=5</td>
<td>K=10</td>
<td></td>
<td>K=5</td>
<td>K=10</td>
<td>K=5</td>
<td>K=10</td>
<td></td>
</tr>
<tr>
<td>MP</td>
<td>0.1531</td>
<td>0.1919</td>
<td>0.2304</td>
<td>0.3511</td>
<td>0.1673</td>
<td>0.2129</td>
<td>0.2509</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MF-BPR</td>
<td>0.4519</td>
<td>0.5001</td>
<td>0.5947</td>
<td>0.7434</td>
<td>0.4376</td>
<td>0.2663</td>
<td>0.3012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fossil</td>
<td>0.4946</td>
<td>0.5354</td>
<td>0.5511</td>
<td>0.7630</td>
<td>0.4746</td>
<td>0.2160</td>
<td>0.2483</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCN</td>
<td>0.5199</td>
<td>0.5726</td>
<td>0.6698</td>
<td>0.8059</td>
<td>0.5090</td>
<td>0.2632</td>
<td>0.3029</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRU4Rec+</td>
<td>0.5346</td>
<td>0.5771</td>
<td>0.6830</td>
<td>0.8136</td>
<td>0.5126</td>
<td>0.2763</td>
<td>0.3169</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPMN</td>
<td>0.5408</td>
<td>0.5883</td>
<td>0.6962</td>
<td>0.8201</td>
<td>0.5245</td>
<td>0.2820</td>
<td>0.3216</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARRec</td>
<td>0.5665</td>
<td>0.6047</td>
<td>0.7102</td>
<td>0.8278</td>
<td>0.5433</td>
<td>0.3009</td>
<td>0.3385</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OAR</td>
<td>0.6078*</td>
<td>0.6415*</td>
<td>0.7425*</td>
<td>0.8462*</td>
<td>0.5847*</td>
<td>0.3200*</td>
<td>0.3580*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* indicates that the improvement of the best result is statistically significant compared with second best result with $p < 0.01$. 
Modeling Global and Personal Occasions

Similarity between different calendar days.

The average predicted preferences

The attention weights by different components
Takeaways

Learnings

• Shopping decisions can be influenced by different occasions, leading to purchases that deviate from a user’s intrinsic preferences. Over Amazon and Etsy, we gain insights into the traceable patterns of personal and global occasion signals.
• We propose to utilize different attention mechanisms to elicit different occasion signals for recommendation. Through experiments, we find the proposed Occasion-Aware Recommender model can outperform the state-of-the-art model in two real-world e-commerce datasets.
• Next, we are interested in introducing more context information to characterize the occasions explicitly and provide explainable recommendations.
Summary of The Talk

Challenge I: Recommendation and Search Eco-System
We need to understand the interplay between recommendation and search modules as well as their whole ecosystem to create a coherent user experience and optimize user engagement.

Challenge II: User Intent Understanding
We need to understand an individual’s shopping needs including short-term, long-term, periodical, impulse and inspirational intents to optimize user engagement.
Thank You