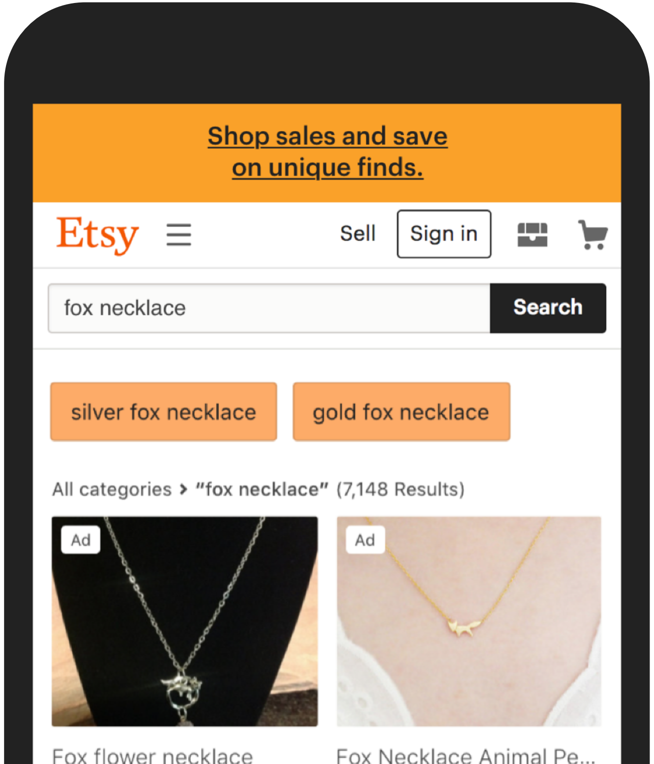
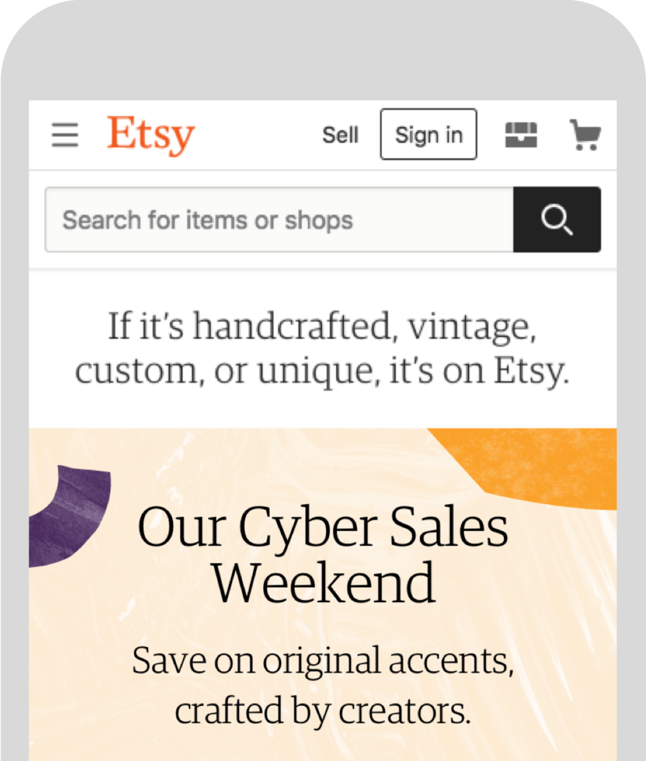
Recent Advances and Challenges in E-Commerce Search & Recommendation Systems

Liangjie Hong
February 3, 2020

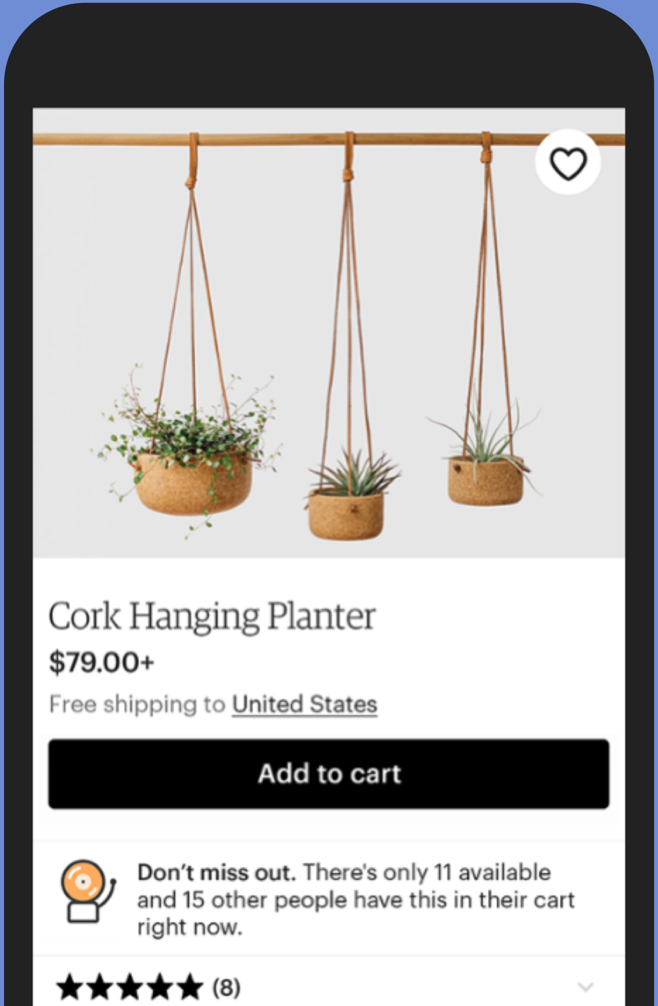
Agenda

1. E-Commerce Search and Recommendation
2. Challenge I:
   Recommendation and Search Eco-System
3. Challenge II:
   User Intent Understanding
4. Summary


Shop sales and save
on unique finds.
Etsy
Sell
Sign in
fox necklace
Search
silver fox necklace
gold fox necklace
All categories › "fox necklace" (7,148 Results)
Ad
Ad
Fox flower necklace
Fox Necklace Animal Pe...


Etsy
Sell
Sign in
Search for items or shops
If it's handcrafted, vintage, custom, or unique, it's on Etsy.
Our Cyber Sales Weekend
Save on original accents, crafted by creators.

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

Data Science and Machine Learning at Etsy

Recent Publications

<u>**Evaluation and Experimentation**</u>

- X. Yin and L. Hong. **The Identification and Estimation of Direct and Indirect Effects in A/B Tests through Causal Mediation Analysis**. In **KDD 2019**.
- N. Ju, D. Hu, A. Henderson and L. Hong. **A Sequential Test for Selecting the Better Variant – Online A/B testing, Adaptive Allocation, and Continuous Monitoring**. In **WSDM 2019**.

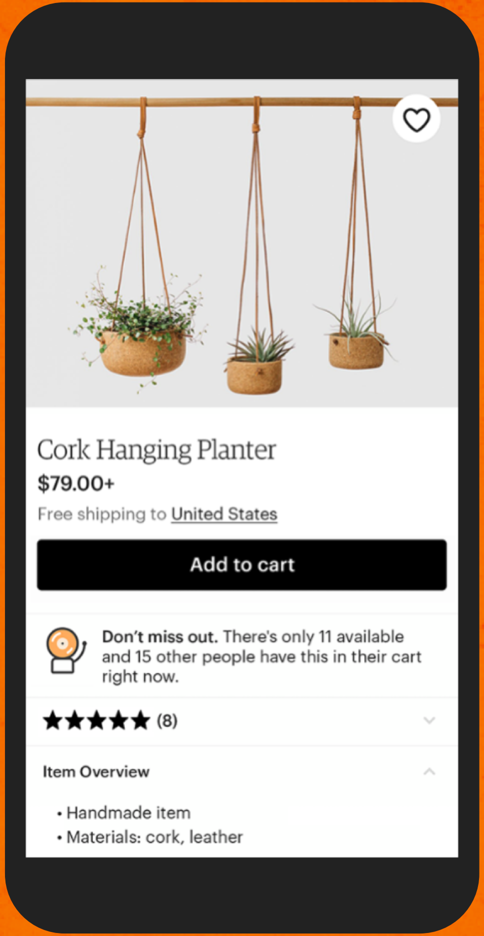
<u>**Search and Recommendation**</u>

- M. Tanjim, C. Su, E. Benjamin, D. Hu, L. Hong and J. McAuley. **Attentive Sequential Model of Latent Intent for Next Item Recommendation**. In **WWW 2020**.
- J. Wang, R. Louca, D. Hu, C. Cellier, J. Caverlee and L. Hong. **Time to Shop for Valentine's Day: Shopping Occasions and Sequential Recommendation in E-commerce**. In **WSDM 2020**.
- R. Louca, M. Bhattacharya, D. Hu and L. Hong. **Joint Optimization of Profit and Relevance for Recommendation Systems in E-commerce**. In the proceedings of RMSE Workshop 2019 at RecSys 2019.
- H. Jiang, A. Sabharwal, A. Henderson, D. Hu and L. Hong. **Understanding the Role of Style in E-commerce Shopping**. In **KDD 2019**.
- A. Stanton, A. Ananthram, C. Su and L. Hong. **Revenue, Relevance, Arbitrage and More: Joint Optimization Framework for Search Experiences in Two-Sided Marketplaces**. ArXiv. 2019.
- X. Zhao, R. Louca, D. Hu and L. Hong. **Learning Item-Interaction Embeddings for User Recommendations**. DAPA at WSDM 2019.
- D. Hu, R. Louca, L. Hong and J. McAuley. **Learning Within-Session Budgets from Browsing Trajectories**. In **RecSys 2018**.
- L. Wu, D. Hu, L. Hong and H. Liu. **Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce**. In **SIGIR 2018**.

<u>**Machine Learning Systems**</u>

- A. Stanton, L. Hong and M. Rajashekhar. **Buzzsaw: A System for High Speed Feature Engineering**. In the proceedings of the 1st SysML Conference.

Challenge I: Recommendation and Search Eco-System



Example I:
An A/B Test Result for A New Recommendation Algorithm

	% Change
Recommendation Clicks	+5%
Search Clicks	-3%
Revenue	~

Example I:
An A/B Test Result for A New Recommendation Algorithm

	% Change
Recommendation Clicks	+5%
Search Clicks	-3%
Revenue	~

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

	% Change
Recommendation Clicks	-10%
Search Clicks	+5%
Revenue	+1%

Example II:
An A/B Test Result for A New Recommendation Algorithm

	% Change
Recommendation Clicks	-10%
Search Clicks	+5%
Revenue	+1%

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.

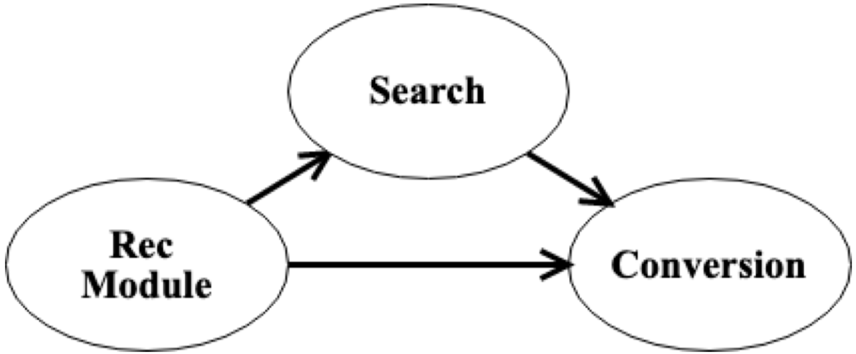
- Opportunity 1:
  Understand experimental results while multiple teams work on different recommendation and search modules.
- Opportunity 2:
  Develop and implement strategies to improve multiple modules and possibly optimize overall user engagement.
- Opportunity 3:
  Develop machine learning models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

X. Yin and L. Hong. **The Identification and Estimation of Direct and Indirect Effects in A/B Tests through Causal Mediation Analysis**. KDD 2019.

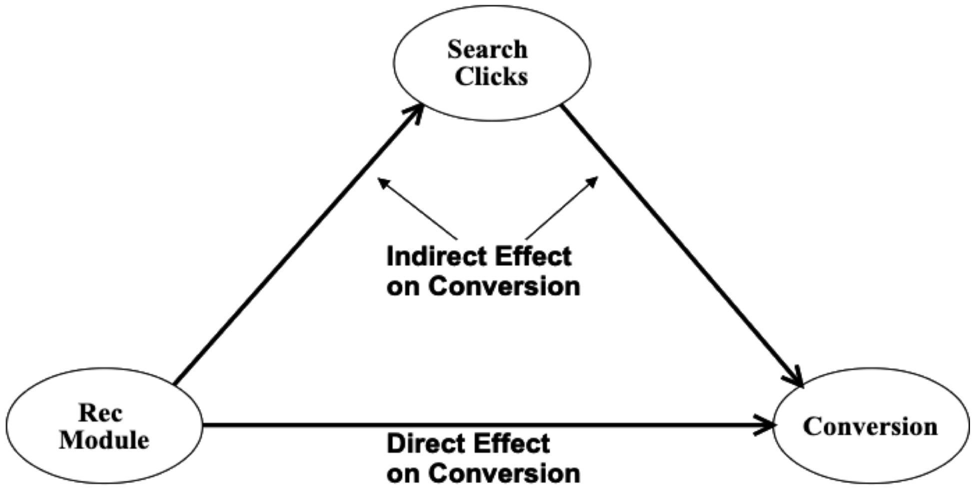
- **Opportunity 1:**
  Understand experimental results while multiple teams work on different recommendation and search modules.
- Opportunity 2:
  Develop and implement strategies to improve multiple modules and possibly optimize overall user engagement.
- Opportunity 3:
  Develop machine learning models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

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

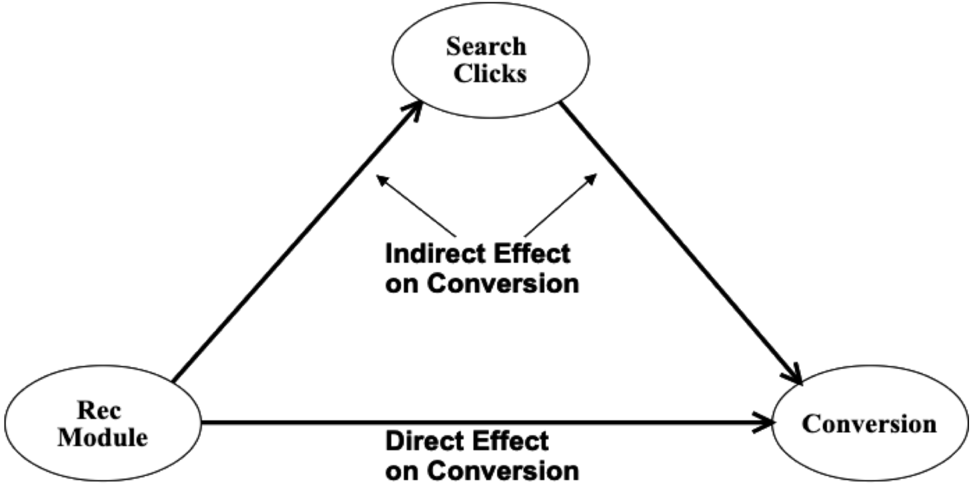


- An improved recommendation module could effectively suggest items that satisfy users' needs so that users don't need to search as much as usual.
- The overall performance of an improved recommendation module could be *cannibalized* by the *induced* reduction of user engagement in search.
- The performance 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.
6. Git Repo: https://github.com/xuanyin/causal-mediation-analysis-for-ab-tests

Case I:
RecSys Listing Page Same-Shop Experiment

	% Change	
Effect	Conversion Rate	GMV
GADE Direct Effect of the Change of Rec Module	**0.4959%***	0.1681%
GACME The Effect of the Induced Change of Search	**-0.2757%*****	**-0.4200%*****
ATE	0.2202%	-0.2518%

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_0: the effect is zero, which is based on asymptotic normality.

Case II:
RecSys Listing Page Internal-Bottom Desktop Experiment

	% Change	
Effect	Conversion Rate	GMV
GADE Direct Effect of the Change of Rec Module	**0.3448%***	0.0659%
GACME The Effect of the Induced Change of Search	**-0.0570%**.	**-0.0926%**.
ATE	**0.2878%**.	-0.0267%

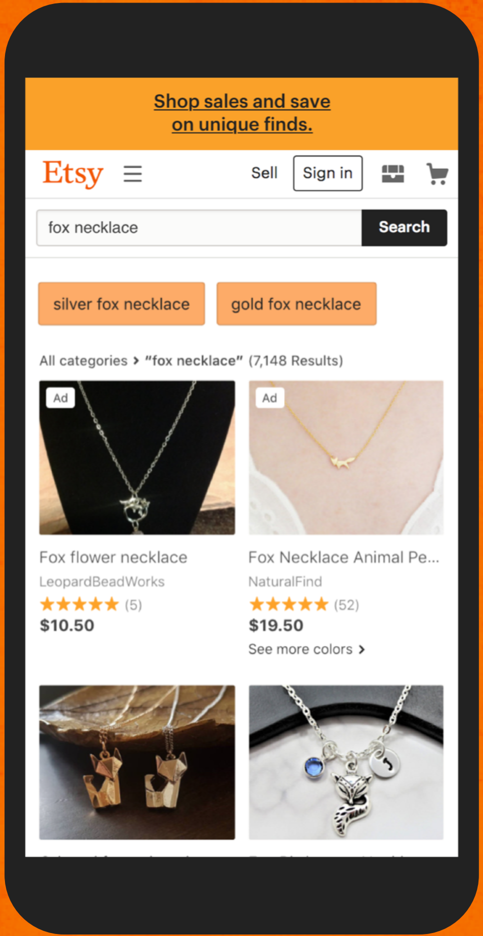
Notes:
1. % Change = Effect/Mean of Control
2. '***' $p<0.001$, '**' $p<0.01$, '*' $p<0.05$, '.' $p<0.1$. Two-tailed p-value is derived from z-test for H_0: the effect is zero, which is based on asymptotic normality.

Takeaways

Learnings

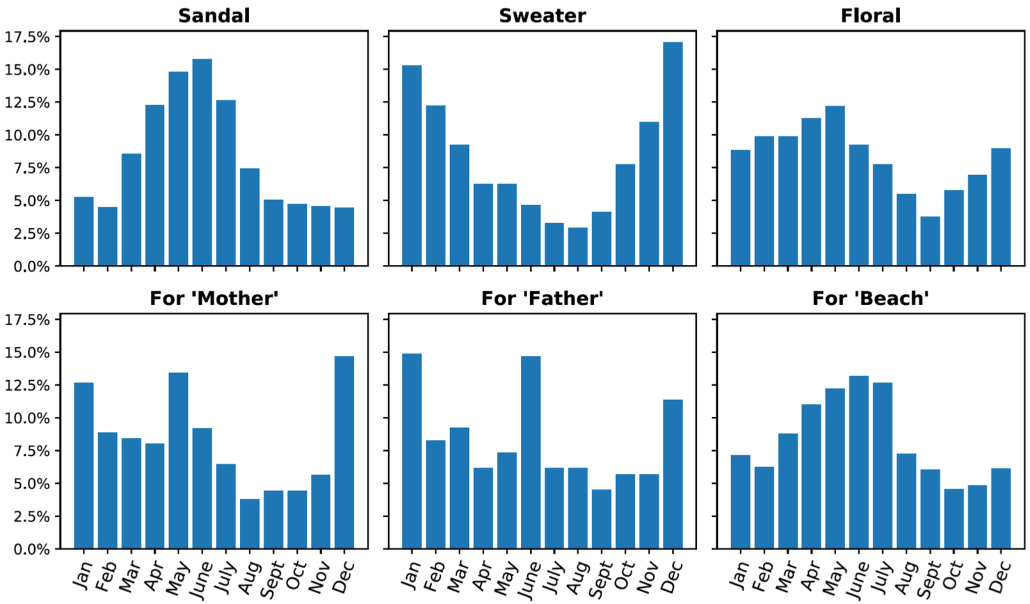
- Understanding direct vs. indirect effects enables us to understand the competition between recommendation modules and search results; and give more informed decisions during roll-outs
- Develop better recommendation strategies such as suggesting items and categories not searched organically or diverse information shown in different surfaces.
- Develop better offline evaluation framework to incorporate both search and recommendation results.

Challenge II: User Intent Understanding



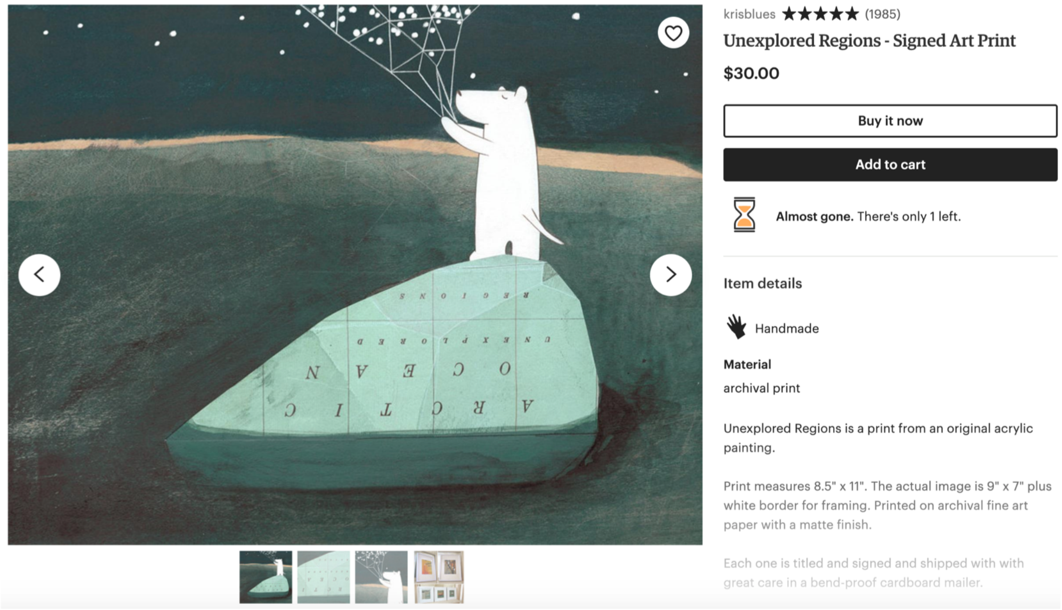
Amazon Users' Shopping Preferences

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 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.

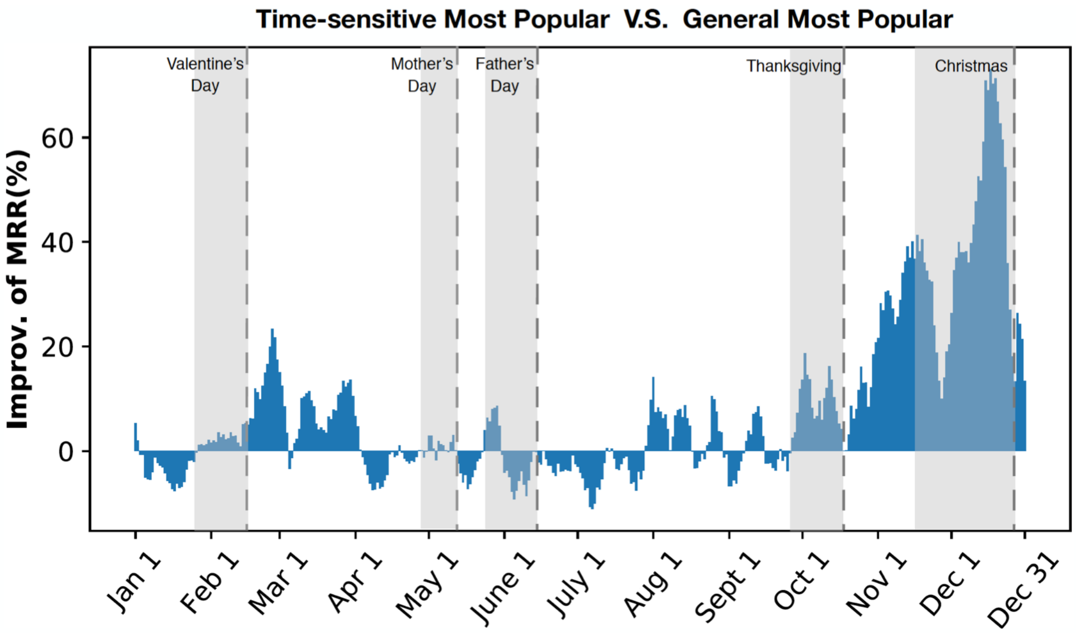
We need to understand an individual's shopping needs including short-term, long-term, periodical, impulse and inspirational intents to optimize user engagement.

J. Wang, R. Louca, D. Hu, C. Cellier, J. Caverlee and **L. Hong**. **Time to Shop for Valentine's Day: Shopping Occasions and Sequential Recommendation in E-commerce**. WSDM 2020.

- **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.

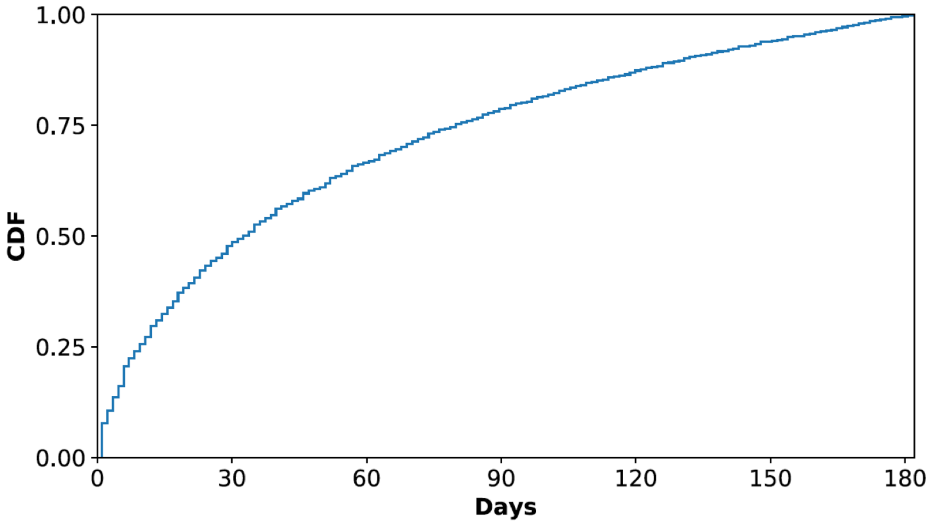
Etsy Users' Shopping Preferences

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



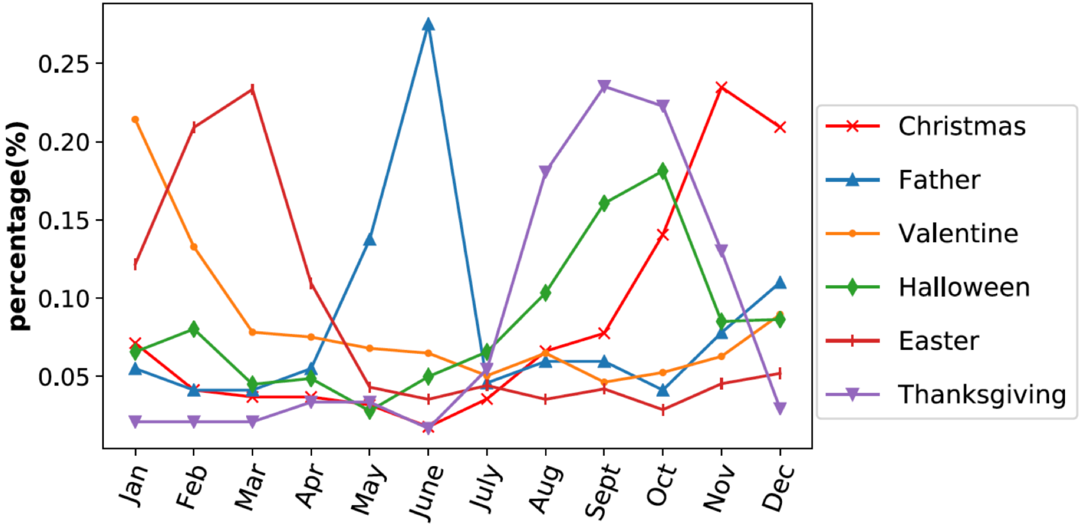
Etsy Users' Shopping Preferences

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.

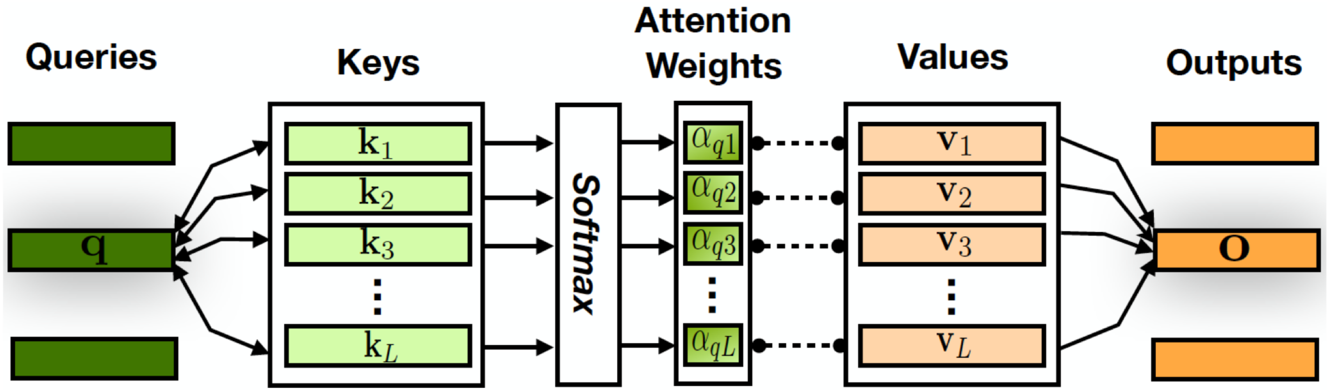


Etsy Users' Shopping Preferences

The reasons an infant's items shopper changes his/her shopping behaviors.



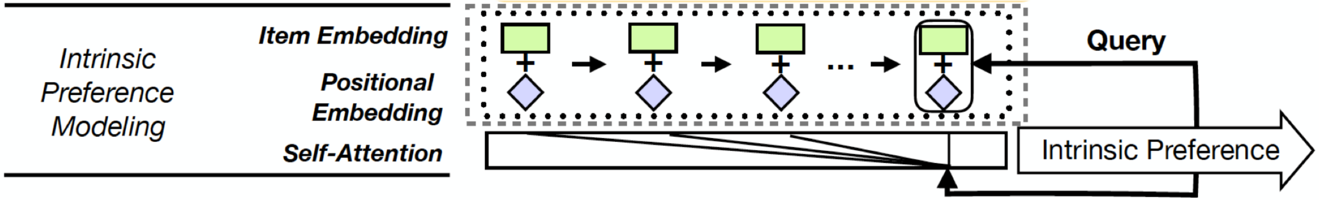
Modeling Global and Personal Occasions



$$\mathbf{o} = \sum_{l=1}^{L} \alpha_{ql}\mathbf{v}_l, \quad \textit{where} \quad \alpha_{ql} = \frac{\exp(s(\mathbf{q}, \mathbf{k}_l))}{\sum_{l=1}^{L} \exp(s(\mathbf{q}, \mathbf{k}_l))}$$

Attention Mechanism

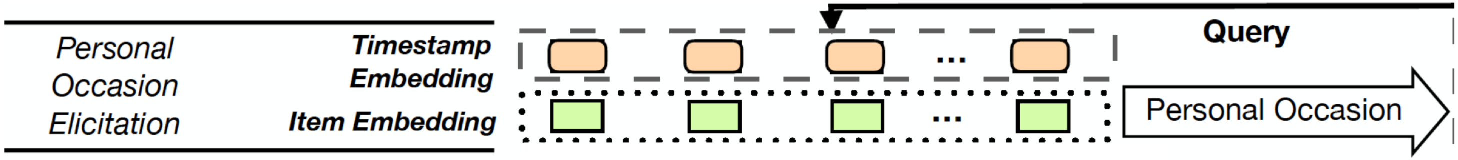
Modeling Global and Personal Occasions



$$\textit{Query} : \hat{\mathbf{m}}^{Q}_{p^u_d} \qquad \textit{Scoring} : s(\mathbf{q}, \mathbf{k}_j) = \frac{\mathbf{q}\mathbf{k}_j^T}{\sqrt{D}}$$

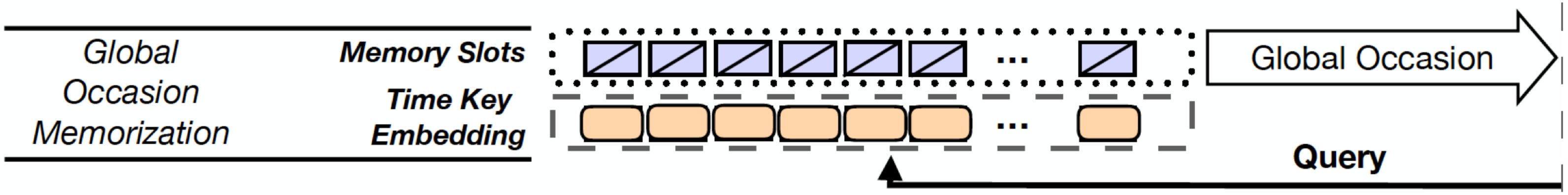
$$\textit{(Key, Value)} : (\hat{\mathbf{m}}^{K}_{p^u_1}, \hat{\mathbf{m}}^{V}_{p^u_1}), (\hat{\mathbf{m}}^{K}_{p^u_2}, \hat{\mathbf{m}}^{V}_{p^u_2}), \ldots, (\hat{\mathbf{m}}^{K}_{p^u_d}, \hat{\mathbf{m}}^{V}_{p^u_d})$$

Modeling Global and Personal Occasions



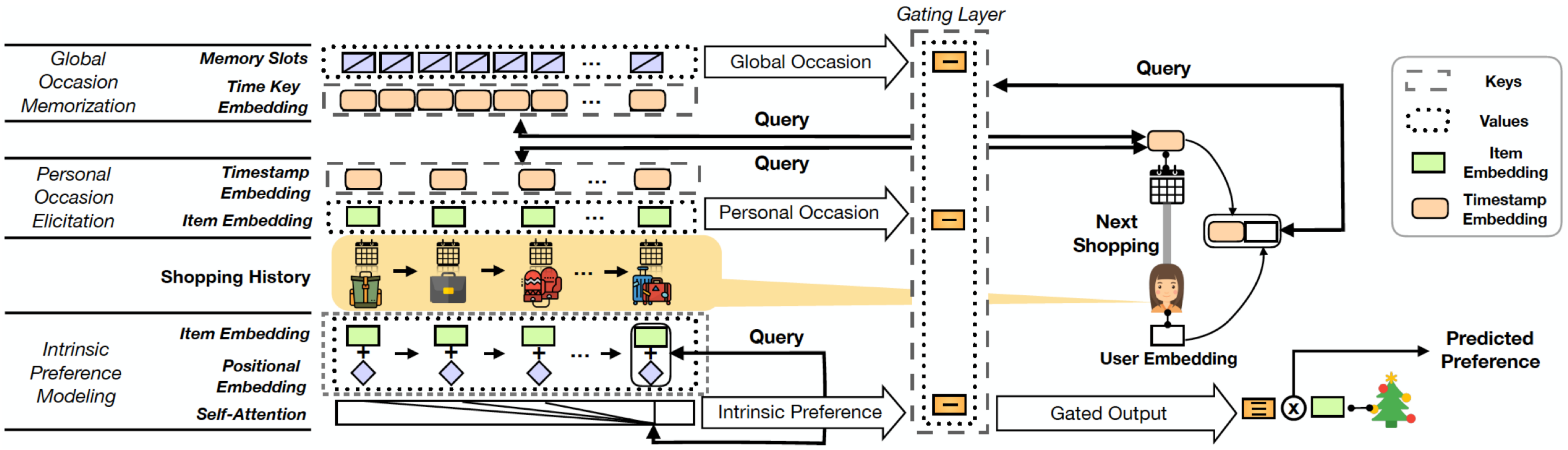
$$Query: \hat{\mathbf{t}}^{Q'}_{t^u_{d+1}} \qquad (Key, Value): (\hat{\mathbf{t}}^{K'}_{t^u_1}, \hat{\mathbf{e}}^{V'}_{p^u_1}), (\hat{\mathbf{t}}^{K'}_{t^u_2}, \hat{\mathbf{e}}^{V'}_{p^u_2}), \ldots, (\hat{\mathbf{t}}^{K'}_{t^u_d}, \hat{\mathbf{e}}^{V'}_{p^u_d})$$

Modeling Global and Personal Occasions



$$\textit{Query}: \hat{\mathbf{t}}^{Q''}_{t^u_{d+1}} \qquad \textit{(Key, Value)}: (\hat{\mathbf{t}}_1, \mathbf{r}_1), (\hat{\mathbf{t}}_2, \mathbf{r}_2), \ldots, (\hat{\mathbf{t}}_M, \mathbf{r}_M)$$

Modeling Global and Personal Occasions



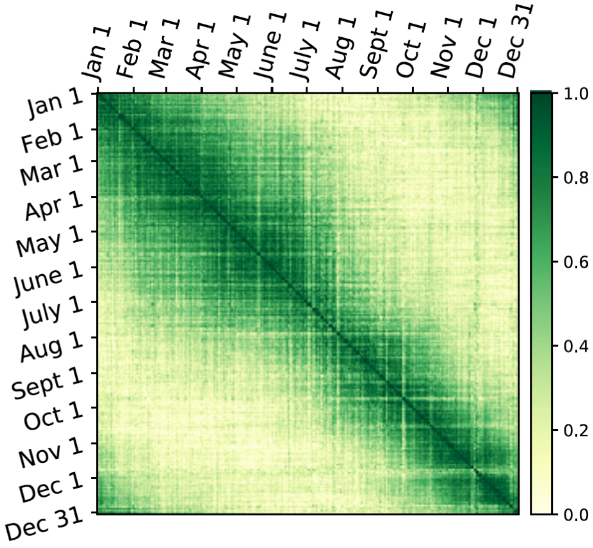
The proposed *Occasion-Aware Recommendation* (OAR) model

Modeling Global and Personal Occasions

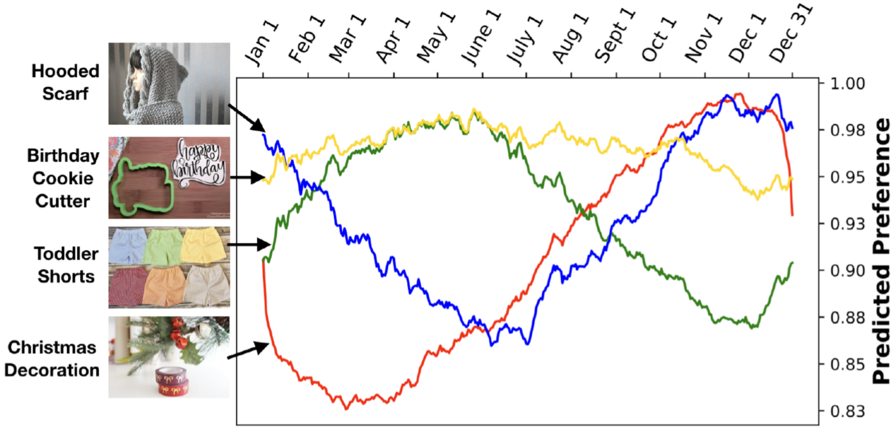
Model	*Etsy*					*Amazon*				
	NDCG		HR		MRR	NDCG		HR		MRR
	K=5	K=10	K=5	K=10		K=5	K=10	K=5	K=10	
MP	0.1531	0.1919	0.2304	0.3511	0.1673	0.2129	0.2509	0.3020	0.4199	0.2195
MF-BPR	0.4519	0.5001	0.5947	0.7434	0.4376	0.2663	0.3012	0.3619	0.4698	0.2668
Fossil	0.4946	0.5354	0.5511	0.7630	0.4746	0.2160	0.2483	0.2967	0.3969	0.2221
TCN	0.5199	0.5726	0.6698	0.8059	0.5090	0.2632	0.3029	0.3664	0.4893	0.2650
GRU4Rec+	0.5346	0.5771	0.6830	0.8136	0.5126	0.2763	0.3169	0.3828	0.5087	0.2770
HPMN	0.5480	0.5883	0.6962	0.8201	0.5245	0.2820	0.3216	0.3881	0.5109	0.2819
SARec	0.5665	0.6047	0.7102	0.8278	0.5433	0.3009	0.3385	0.4085	0.5251	0.2984
OAR	**0.6078***	**0.6415***	**0.7425***	**0.8462***	**0.5847***	**0.3200***	**0.3580***	**0.4301***	**0.5476***	**0.3165***

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

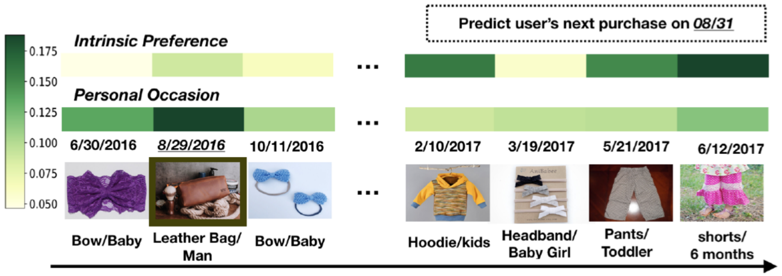
Modeling Global and Personal Occasions



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