Style in the Long Tail

Discovering Unique Interests with Latent Variable Models in Large Scale Social E-commerce

Diane Hu, Etsy
Rob Hall, Etsy
Josh Attenberg, Etsy
Overview

01 | Etsy Overview

02 | Discovering User Styles

03 | Generating Recommendations
Etsy is an online marketplace where people connect to buy and sell **unique** goods: Handmade, vintage, or craft supplies.
Etsy is an online marketplace where people connect to buy and sell unique goods: Handmade, vintage, or craft supplies.

- 40M Users
- 1M Shops
- 26M Listings
- 9 Languages
- 200 Countries

How to build recommender systems for such a unique marketplace?
How do people typically use Etsy?

1. **Browse (Unintentional)**
   - via the front page, browse pages or the activity feed

2. **Favorite**
   - an item or shop, and add to collections with coherent theme/style

3. **Follow**
   - another user with similar style/interest

**Typical E-Commerce Usage**
How do people decide what to buy?

Function and style. Example: search results for “nightstand” - 100+ pages
How to describe style?
What Personalization Looks Like on Etsy

Recommendations for **multiple intents**
Enhance browsing experience, not just purchases

Recommendations for **multiple content types**
Develop unified method for recommending shops, items, users

Recommendations based on **visual styles**
Identify user styles and interests in a visually transparent way
Overview

01 | Etsy Overview
02 | Discovering User Styles
03 | Generating Recommendations
Solution Overview

1. Learn **style profiles** for each user using LDA

   - 10% "surreal"
   - 60% "geometric"
   - 30% "mid-century modern"

2. Define what each style looks like

   - "mid-century modern"

3. User style profiles to generate personalized content

   - ITEM RECS
   - USER REC
   - SHOP REC
Assume: Each user’s favorited items are generated by this process:

For each user $u$,

1. Draw a style profile:
   $\theta \sim \text{Dirichlet}(\alpha)$

2. For each item, $x_n$ that user $u$ has favorited,
   
   (a) Draw a style:
       $z_n \sim \text{Multinomial}(\theta)$
   
   (b) Draw an item:
       $x_n \sim \text{Multinomial}(\beta_{z_n})$
Latent Dirichlet Allocation (LDA) for Discovering Styles

Discover popular styles on Etsy as a distribution over items

*“geometric” “mid-century” “surreal”*

Represent each user as a distribution over popular styles, i.e. “style profile”

\[
\begin{align*}
&\text{Diane Hu} \\
&\text{Brooklyn, NY} \\
&0.38 \\
&0.13 \\
&0.02 \\
&\cdots
\end{align*}
\]

“geometric” “mid-century” “surreal”
DISCOVERING USER STYLES

Different styles discovered by LDA

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Botanical" /></td>
<td><img src="image" alt="Surreal landscapes" /></td>
<td><img src="image" alt="Whimsical" /></td>
<td><img src="image" alt="Acrylic/Abstract" /></td>
<td><img src="image" alt="French Dolls" /></td>
<td><img src="image" alt="Whimsical/Abstract" /></td>
<td><img src="image" alt="Cities" /></td>
<td><img src="image" alt="Vintage" /></td>
</tr>
</tbody>
</table>

Example of learned styles that contain art prints:

- A = Botanical
- B = Surreal landscapes
- C = Whimsical
- D = Acrylic/Abstract
- E = French Dolls
- F = Whimsical/Abstract
- G = Cities
- H = Vintage
DISCOVERING USER STYLES

Different styles discovered by LDA

ANIMALS

GEOMETRIC

TENTACLES

MID-CENTURY MODERN
Overview

01 | Etsy Overview

02 | Discovering User Styles

03 | Generating Recommendations
User Recommendations

Given that each user has a style profile: **Recommend N users with most similar style profiles.**

**Brute-force K-NN is too expensive.** Hash similar users into the same bins, and perform K-NN within each bin.

- **Locality Sensitive Hashing (LSH).** Create hash based on which side of a series of random planes the user falls onto.

- **“Top-K” Hashing.** Create hash based on set of all pairs of top-k topic indices.
Item Recommendations

Given that each user has a style profile: Sample items most highly weighted styles

USER'S FAVORITES

ITEM RECOMMENDATIONS

STYLE #428

STYLE #655

STYLE #54

STYLE #87
Shop Recommendations

Re-learn topic models substituting item ids with shop ids. Sample shops from highly weighted styles.
Shop Recommendations

Re-learn topic models substituting item ids with shop ids. Sample shops from highly weighted styles.
Recommending Styles within Categories

Find how overall styles translate into specific categories
Visualizing Related Topics

Learn topic correlations from users’ style-profiles.

“Bright, Whimsical”
Visualizing Related Topics

Learn topic correlations from users’ style-profiles.

“Gothic Punk”
Visualizing Related Topics

Learn topic correlations from users’ style-profiles.

“Sci-fi/Fantasy”
Recommendations in the Activity Feed

ORGANIC

ITEM REC

SHOP REC

USER REC
User Recommendation Experiments

Side-by-Side User Study

- Randomly interleave user recs from 3 algorithms: (1) LDA, (2) TF-IDF w/ Cosine Similarity, (3) Triadic Closure
- User rated each recommendation positive, neutral, negative
- LDA was overwhelming winner

A/B Testing in Activity Feed

*Phase One:*
- LDA vs. No recs
- Significantly increased all business metrics

*Phase Two:*
- Different variants of LDA vs. Matrix Factorization (using Stochastic SVD)
- Matrix factorization and LDA comparable across business metrics
Conclusion

What We Did

• Identify styles across Etsy as a visual experience

• Generate style profiles that are visually transparent and capture diverse taste

• Build large-scale recommender systems:
  ‣ for multiple content types
  ‣ for enhancing browse experience

• Improve key business metrics

More Details On

• System/hard-ware set-up

• Scaling algorithms to ~40M users

• Experimental set-up and outcomes

• Product design for recommendations