Visually-Aware Fashion Recommendation and Design with Generative Image Models

Wang-Cheng Kang¹, Chen Fang², Zhaowen Wang² & Julian McAuley¹
UC San Diego¹ & Adobe Research²
ICDM’17, New Orleans
Outlines

- **Introduction to fashion**
  - Who is interested in fashion?
  - Characteristics of fashion items

- **Personalized Fashion recommendation and design**
  - Needs and challenges
  - Methodology
  - Evaluation

- **Conclusion and future work**
Who is interested in fashion?
People always have great interest in fashion.
Fashion companies
E-commercial sites

Amazon
Amazon Fashion
Asos Fashion Finder
Ebay Fashin
Polyvore
Taobao Fashion
E-commercial sites

Alibaba’s AI Fashion Consultant Helps Achieve Record-Setting Sales

AI will blur the line between online and offline retail.

by Yiting Sun  November 13, 2017

Amazon Has Developed an AI Fashion Designer

The retail giant is taking a characteristically algorithmic approach to fashion.

by Will Knight  August 24, 2017
Our work was recently featured on MIT Tech Review, Daily Mail, Yahoo Finance, The Merkle, Fossbytes…
Researchers

- “Machine learning meets fashion” workshop @ KDD’16,’17
Characteristics of fashion items
Large in amount

- Amazon.com

Number of products in each category:

- No1. Books **2.3 M**
- No2. Clothing, Shoes and Jewelry **1.5M**
- No3. Sports and Outdoors **0.5M**
- ...
Diverse and Subtle in style
Volatile in defining what is fashionable

- Fashion is **subjective** and changes through time

[He and McAuley, WWW 2016]
Needs in fashion domain

- Consumers: Can I easily find favorite fashion items?
- Online Sellers: Can I recommend items more accurately?

- Fashion companies:
  - Can I evaluate newly designed products?
  - Can I get inspiration when designing new fashion items?
  - Can I know why the consumer doesn’t like the clothing?
Needs in fashion domain

- Consumers: Can I easily find favorite fashion items?
- Online Sellers: Can I recommend items more accurately?
- Fashion companies:
  - Can I evaluate newly designed products?
  - Can I get inspiration when designing new fashion items?
  - Can I know why the consumer doesn't like the clothing?
Challenges in Fashion Recommendation and Design

- Visual signal is important in this domain
  - Usually ignored in traditional recsys, due to the lack of datasets
  - Style is hard to capture due to its complexity
- Cold start problem
  - Long tail effect *(many are cold)*
  - New clothing are constantly introduced *(all are cold)*
  - Need to design unseen clothing *(completely cold)*
- How to generate unseen clothing
- How to link content recommendation and content generation
Preference predictor

- Biased Matrix Factorization
  \[ x_{ui} = \alpha + \beta_u + \beta_i + \gamma_u^T \gamma_i \]

- VBPR (Visual Bayesian Personalized Ranking)
  \[ x_{ui} = \alpha + \beta_u + \beta_i + \gamma_u^T \gamma_i + \theta_u^T (E_{f_i}) \]

- DVBPR
  - Joint learning style embedding with CNN
  \[ x_{ui} = \alpha + \beta_u + \beta_i + \gamma_u^T \gamma_i + \theta_u^T (E_{f_i}) \]
Preference predictor

- Biased Matrix Factorization
  \[ x_{ui} = \alpha + \beta_u + \beta_i + \gamma_u^T \gamma_i \]

- VBPR (Visual Bayesian Personalized Ranking)
  \[ x_{ui} = \alpha + \beta_u + \gamma_i^T \gamma_i + \theta_u^T (Ef_i) \]

- DVBPR
  - Joint learning style embedding with CNN
  \[ x_{ui} = \alpha + \beta_u + \beta_i + \gamma_u^T \gamma_i + \theta_u^T \Phi(Img_i) \]
Preference predictor

- Biased Matrix Factorization
  \[ x_{ui} = \alpha + \beta_u + \beta_i + \gamma_u^T\gamma_i \]

- VBPR (Visual Bayesian Personalized Ranking)
  \[ x_{ui} = \alpha + \beta_u + \beta_i + \gamma_u^T\gamma_i + \theta_u^T(E_f_i) \]

- DVBPR
  - Joint learning style embedding with CNN
  - Remove item bias terms for adapting to new samples
  \[ x_{ui} = \alpha + \beta_u + \beta_i + \gamma_u^T\gamma_i + \theta_u^T\Phi(Img_i) \]
Preference predictor

- Biased Matrix Factorization
  \[ x_{ui} = \alpha + \beta_u + \beta_i + \gamma_i^T \gamma_i \]

- VBPR (Visual Bayesian Personalized Ranking)
  \[ x_{ui} = \alpha + \beta_u + \beta_i + \gamma_i^T \gamma_i + \theta_u^T (E f_i) \]

- DVBPR
  - Joint learning style embedding with CNN
  - Remove item bias terms for adapting to new samples
  \[ x_{ui} = \alpha + \beta_u + \theta_u^T \Phi(I m g_i) \]
Preference predictor

- Biased Matrix Factorization
  \[ x_{ui} = \alpha + \beta_u + \beta_i + \gamma_u^T \gamma_i \]

- VBPR (Visual Bayesian Personalized Ranking)
  \[ x_{ui} = \alpha + \beta_u + \beta_i + \gamma_u^T \gamma_i + \theta_u^T (Ef_i) \]

- DVBPR
  - Joint learning style embedding with CNN
  - Remove item bias terms for adapting to new samples
  - Cancel global bias and user bias for optimizing ranking
  \[ x_{ui} = \alpha + \beta_u + \theta_u^T \Phi(Img_i) \]
Preference predictor

- Biased Matrix Factorization
  \[ x_{ui} = \beta_i + \gamma_u^T \gamma_i \]
- VBPR (Visual Bayesian Personalized Ranking)
  \[ x_{ui} = \beta_i + \gamma_u^T \gamma_i + \theta_u^T (E f_i) \]
- DVBPR
  - Joint learning style embedding with CNN
  - Remove item bias terms for adapting to new samples
  - Cancel global bias and user bias for optimizing ranking
  \[ x_{ui} = \theta_u^T \Phi(I m g_i) \]
GANs

- Effectively approximate a distribution
- Efficiently generate samples from a **continuous space**

- Generator $G(z,c)$
  - Take a noisy $z$ and a category $c$ as inputs
  - Output a generated image

- Discriminator $D(x,c)$
  - Judge whether the input image $x$ is real (from dataset) or fake (from $G$)
Training Process

- Optimization objective
  - **BPR** (Bayesian personalized ranking) for implicit feedback
  - Approximately optimize the AUC of personalized ranking
  - For each user, positive items should be ranked higher than other items

- Training Algorithm
  - **SGD** (stochastic gradient descent)
  - Joint learning with a (siamese) CNN

- We also train a conditional GAN on the fashion dataset
  - The GAN can generate a new fashion item given a category $c$ and a low-dimensional vector $z$
Preference Maximization

- Find an item that is most favored by a given user
- **Retrieval-based method**

\[
\delta(u, c) = \arg\max_{e \in X_c} x_{u,e} = \arg\max_{e \in X_c} \theta_u^T \Phi(e),
\]

- **Generation-based method**

\[
\hat{\delta}(u, c) = \arg\max_{e \in G(\cdot, c)} \hat{x}_{u,e} = \arg\max_{e \in G(\cdot, c)} x_{u,e} - \eta L_{real}(e, c) \]

\[
= G \left[ \arg\max_{z \in [-1,1]^{100}} \theta_u^T \Phi[\Delta G_c(z)] - \eta[D_c(G_c(z)) - 1]^2, c \right],
\]
Preference Maximization

- Find an item that is most favored by a given user

**Retrieval-based method**

\[
\delta(u,c) = \arg\max_{e \in X_c} x_{u,e} = \arg\max_{e \in X_c} \theta^T \Phi(e),
\]

**Generation-based method**

\[
\overset{\hat{}}{\delta}(u,c) = \arg\max_{e \in G(\cdot,c)} \hat{x}_{u,e} = \arg\max_{e \in G(\cdot,c)} x_{u,e} - \eta L_{real}(e,c) = G\left[ \arg\max_{z \in [-1,1]}^{100} \theta_u^T \Phi[\Delta_G(z)] - \eta [D_G(G_c(z)) - 1]^2, c \right],
\]

Retrieve in dataset

Trade-off between Pref. score and quality

Optimize in “GAN space”
Datasets

### Dataset Statistics (After Preprocessing)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Users</th>
<th>#Items</th>
<th>#Interactions</th>
<th>#Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon Fashion</td>
<td>64583</td>
<td>234892</td>
<td>513367</td>
<td>6</td>
</tr>
<tr>
<td>Amazon Women</td>
<td>97678</td>
<td>347591</td>
<td>827678</td>
<td>53</td>
</tr>
<tr>
<td>Amazon Men</td>
<td>34244</td>
<td>110636</td>
<td>254870</td>
<td>50</td>
</tr>
<tr>
<td>Tradesy.com</td>
<td>33864</td>
<td>326393</td>
<td>655409</td>
<td>N/A</td>
</tr>
</tbody>
</table>

- Implicit feedback
- Each item associates with an image
- **1M** items, **230K** users and **2M** interactions in total (highly sparse)
Evaluation Metrics

- Recommendation
  - AUC on all items and cold items

- Retrieval vs. Generation
  - Preference score: mean preference score
  - Image quality: inception score
  - Image diversity: opposite SSIM

- Image metrics are limited and subjective, we also provide qualitative results
# Recommendation performance

## Content-unaware vs. Content-aware

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Setting</th>
<th>(a) RAND</th>
<th>(b) PopRank</th>
<th>(c) WARP</th>
<th>(d) BPR-MF</th>
<th>(e) VisRank</th>
<th>(f) FM</th>
<th>(g) VBPR</th>
<th>(h) DVBPR</th>
<th>Improvement h vs. d</th>
<th>Improvement h vs. best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon Fashion</td>
<td>All Items</td>
<td>0.5</td>
<td>0.5849</td>
<td>0.6065</td>
<td>0.6278</td>
<td>0.6839</td>
<td>0.7093</td>
<td>0.7479</td>
<td><strong>0.7964</strong></td>
<td>26.9%</td>
<td>6.5%</td>
</tr>
<tr>
<td></td>
<td>Cold Items</td>
<td>0.5</td>
<td>0.3905</td>
<td>0.5030</td>
<td>0.5514</td>
<td>0.6807</td>
<td>0.7088</td>
<td>0.7319</td>
<td><strong>0.7718</strong></td>
<td>40.0%</td>
<td>5.5%</td>
</tr>
<tr>
<td>Amazon Women</td>
<td>All Items</td>
<td>0.5</td>
<td>0.6192</td>
<td>0.5998</td>
<td>0.6543</td>
<td>0.6512</td>
<td>0.6678</td>
<td>0.7081</td>
<td><strong>0.7574</strong></td>
<td>15.8%</td>
<td>7.0%</td>
</tr>
<tr>
<td></td>
<td>Cold Items</td>
<td>0.5</td>
<td>0.3822</td>
<td>0.5017</td>
<td>0.5196</td>
<td>0.6387</td>
<td>0.6682</td>
<td>0.6885</td>
<td><strong>0.7137</strong></td>
<td>37.4%</td>
<td>3.7%</td>
</tr>
<tr>
<td>Amazon Men</td>
<td>All Items</td>
<td>0.5</td>
<td>0.6060</td>
<td>0.6081</td>
<td>0.6450</td>
<td>0.6589</td>
<td>0.6654</td>
<td>0.7089</td>
<td><strong>0.7410</strong></td>
<td>14.9%</td>
<td>4.5%</td>
</tr>
<tr>
<td></td>
<td>Cold Items</td>
<td>0.5</td>
<td>0.3903</td>
<td>0.5005</td>
<td>0.5132</td>
<td>0.6545</td>
<td>0.6705</td>
<td>0.6863</td>
<td><strong>0.6923</strong></td>
<td>34.9%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Tradesy.com</td>
<td>All Items</td>
<td>0.5</td>
<td>0.5268</td>
<td>0.6176</td>
<td>0.5860</td>
<td>0.6457</td>
<td>0.7662</td>
<td>0.7500</td>
<td><strong>0.7857</strong></td>
<td>34.1%</td>
<td>2.5%</td>
</tr>
<tr>
<td></td>
<td>Cold Items</td>
<td>0.5</td>
<td>0.3946</td>
<td>0.5333</td>
<td>0.5418</td>
<td>0.6084</td>
<td>0.7730</td>
<td>0.7525</td>
<td><strong>0.7793</strong></td>
<td>43.8%</td>
<td>0.8%</td>
</tr>
</tbody>
</table>
Retrieval vs Generation

Comparison of Top-3 Returned Images for a Given User and Category (1000 Trials). For all three metrics, a larger value is better. The value after the ± is the standard deviation.

<table>
<thead>
<tr>
<th>Item Source</th>
<th>Preference Score</th>
<th>Quality</th>
<th>Diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>6.8685±.36</td>
<td>0.5585±.09</td>
</tr>
<tr>
<td>Random Methods</td>
<td>-1.9547±3.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_c$</td>
<td>-1.9392±3.7</td>
<td>6.8121±.37</td>
<td>0.5589±.09</td>
</tr>
<tr>
<td>$G(\cdot, c)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personalized Methods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta(u, c)$ in (eq. 8)</td>
<td>7.1876±3.2</td>
<td>7.6492±.27</td>
<td>0.5393±.11</td>
</tr>
<tr>
<td>$\hat{\delta}(u, c)$ in (eq. 9)</td>
<td>7.6760±4.1</td>
<td>7.6524±.24</td>
<td>0.5265±.12</td>
</tr>
</tbody>
</table>

Effect of Preference Maximization

Better preference score
Similar Quality
Slightly worse diversity
Effect of the hyper-parameter

\[ \hat{\delta}(u, c) = \arg\max_{e \in G(u, c)} \hat{x}_{u,e} = \arg\max_{e \in G(u, c)} x_{u,e} - \eta L_{real}(e, c) \]
Recommend exiting items and design new items for the user

Generated Clothing

- Higher preference score
- Similar in style but still different

Getting inspirations when designing new items for a (group of) user(s)
Knowing what consumers want beyond items that exist

The user likes the T-shirt with another color

Modifications Types
- Changing color
- Extending sleeve
- ‘distressing’ pants
- shortening pants
- Other minor stylistic changes
Needs in fashion domain

- Consumers: Can I easily find favorite fashion items?
- Online Sellers: Can I recommend items more accurately?

Fashion companies:
- Can I evaluate newly designed products?
- Can I get inspiration when designing new fashion items?
- Can I know why the consumer doesn’t like the clothing?
Needs in fashion domain

- Consumers: Can I easily find favorite fashion items?
- Online Sellers: Can I recommend items more accurately?
- Fashion companies:
  - Can I evaluate newly designed products?
  - Can I get inspiration when designing new fashion items?
  - Can I know why the consumer doesn’t like the clothing?
Conclusion:

- Personalized Fashion Recommendation
- Personalized Fashion Design
- Purple Dress...

Future work:

- Joint training with generative models to improve performance
- Fine-grained style control in generation
- Modeling fashion evolution
- Discovering taste for a user group
- Outfit design
- Exploring other forms of generation
- Gathering richer datasets
- More evaluation methods
Welcome to the fashion world!

Thanks for listening! Any questions?
Generated samples vs real samples

- Similar but different
Interpolation

(a) Discrete Real Image Space in Dataset

(b) Continuous Synthetic Image Space in GAN
Preference maximization (64*64 resolution)