Fashionpedia-Taste
A Dataset towards Explaining Human Fashion Taste
Shi, Mengyun; Belongie, Serge; Cardie, Claire

Publication date: 2023

Document version
Publisher's PDF, also known as Version of record

Document license: CC BY

Citation for published version (APA):
Fashionpedia-Taste: A Dataset towards Explaining Human Fashion Taste

Mengyun Shi¹  Serge Belongie²  Claire Cardie¹
¹Cornell University  ²University of Copenhagen

Abstract

Existing fashion datasets don’t consider the multi-facts that cause a consumer to like or dislike a fashion image. Even two consumers like a same fashion image, they could like this image for totally different reasons. In this paper, we study the reason why a consumer like a certain fashion image. Towards this goal, we introduce an interpretability dataset, Fashionpedia-taste, consist of rich annotation to explain why a subject like or dislike a fashion image from the following 3 perspectives: localized attributes, human attention, and caption. Furthermore, subjects are asked to provide their personal attributes and preference on fashion, such as personality and preferred fashion brands. Our dataset makes it possible for researchers to build computational models to fully understand and interpret consumers’ fashion taste from different humanistic perspectives and modalities. Fashionpedia-taste is available at: ¹

1. Introduction

Why users click ‘like’ on a fashion image shown on social platforms? They use ‘like’ to express their preference for fashion. In the context of fashion, apparel represents second highest e-commerce shopping category. Therefore, fully understand users’ fashion taste could play an important role on the fashion E-commerce.

To increase the chance for a consumer to buy fashion products, a various of recommendation systems have been developed and they have delivered decent results. However, even a recommendation system makes a correct recommendation for a consumer, does that mean the model really understand the reason why this user like this image? The an-

¹Fashionpedia project page: fashionpedia.github.io/home/
2. Related Work

Fashion dataset Most of the previous fashion datasets focus on recognition, detection, data mining, or retrieval tasks (Table 1). In the domain of interaction between users and fashion images, Fashion IQ [26] provides human-generated captions that distinguish similar pair of garment images through natural language feedback. ViBE [8] introduces a dataset to understand users’ fashion preference based on her specific body shape. Fashionpedia-Ads studied the correlation between ads and fashion taste among users. Unlike Fashion IQ, ViBE and Fashionpedia-Ads, our dataset focuses on explaining and reasoning on users’ fashion preference based on both visual and textual signal. Beyond fashion domain, the most relevant work is VCR [31], which requires models to answer correctly and then provide a rationale justification to its answer. Unlike VCR, our dataset requires models to complete more complicated multi-stage reasoning through different modalities (task1/2/3/4) for subjects’ fashion taste, as illustrated in Fig. 1.

3. Fashion Taste Annotation from Subjects

Subjects and Annotation pipeline We recruited 100 female subjects from a U.S. university. Our user annotation process consists of 2 parts: 1) collect subjects’ basic information (Sec. 3.1); 2) collect subjects’ fashion taste for given dress length categories (Sec. 3.2). All the subjects are required to complete these two processes.

3.1. Basic Information Survey

This survey is used to collect the subjects’ basic information (gender, ethnicity, age); 2) personality; 3) basic fashion preference (favorite fashion brands, fashion attributes and categories); 4) favorite dress length, which is used to determine the images from which dress length should be assigned to each subject for the survey mentioned in Sec. 3.2.

Personality Similar to [18], we use the 10-item multiple-choice questions to measure subjects’ personality. We collect personality data because we want to see whether there is a correlation between subjects’ personality and their fashion taste.

Basic fashion preference We collect users’ fashion preference (fashion categories, fashion attributes, and brands) because we are curious whether this self-reported fashion preference is aligned with their preference measured in Sec. 3.2.

3.2. Fashion Taste Survey

Task Design In the fashion taste survey, the subjects are given 100 dress images based on their favourite dress lengths reported in their basic information survey (Sec. 3.1). They are required to tell whether they like these dresses and provide the reasons why they like or dislike these dresses. For each given dress image, they need to complete the following 4 tasks:

- Task 1: judge whether they like or dislike a given dress.
- Task 2: Attribute selection: explain which aspects make them like and dislike a given dress.

<table>
<thead>
<tr>
<th>Task Design</th>
<th>Dataset name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognition</td>
<td>iMat [3], Deep [15], Clothing [29], F-MNIST [27], F-550k [11], F-128 [21], F-14 [23], Hipster [14]</td>
</tr>
<tr>
<td>Detection</td>
<td>ModaNet [32], Deep2 [2], Main [19]</td>
</tr>
<tr>
<td>Data Mining</td>
<td>Vintage [9], Chic [28], Ups [5], Latent [6], Geo [16]</td>
</tr>
<tr>
<td>Retrieval</td>
<td>DARN [10], WTBI [13], Zappos [39], Deep [15]</td>
</tr>
<tr>
<td>Attribute</td>
<td>Capsule [7], POG [1], VIBE [8], IQ [26]</td>
</tr>
<tr>
<td>Exploreability</td>
<td>Fashionpedia-Taste</td>
</tr>
</tbody>
</table>
• Task 3-Human attention: indicate (draw polygons) the regions of the dress that make them like and dislike a given dress.

• Task 4-Textual explanation: explain why the regions they draw from task 3 make them like and dislike a given dress.

**Why we design these 4 tasks** Task 2, 3 and 4 allow the subjects to explain their fashion taste from 3 different perspectives and modalities. Task 2 allows the subjects to explain their fashion taste on the perspective of fine-grained attributes. However, Task 2 might miss to capture some information that can only be explained visually. Task 3 is used to address this issue and simulates human gaze capture, allowing the subjects to explain their fashion taste visually. Furthermore, to fully understand the area that subjects draw in task 3, we use Task 4 to allow the subjects to further explain why their draw the areas in Task 3 textually.

**Imbalanced likes and dislikes**: we expect it will have big data imbalance if we only ask the subjects to explain the reasons that make them like a dress. To address this issue, we asked users to explain both the aspects that make them like and dislike a given image for task 2, 3, and 4.

4. Dataset Analysis

4.1. User annotation analysis

**User basic info**

Task1-Like / dislike We collect 4766 likes and 5234 dislikes, over 1500 unique images, and 100 unique users. The frequency of likes and dislikes selected by each user is shown in Fig. 2, indicating most of the users have fairly balanced like and dislike ratio. Balanced like and dislike ratio could potentially help train less biased models.

Task2-Attribute selection Table 2 breaks down the frequency of liked and disliked attributes selected by each user into different dress lengths. The average liked attributes (3.9397) is nearly 3 times more than disliked attributes (1.4565). This indicates most of the users tend to select the attributes that they like rather than dislike while explaining why they like/dislike a dress.

Table 3 displays the details of annotated attributes into their corresponding fine-grained categories (super-categories). 'Silhouette' contains highest percentage of liked attributes (21.9 %). This indicates most of users’ fashion preference is determined by the shape of a dress. In contrast, 'textile finishing and manufacturing technique' (Tex fini, manu-tech) contains highest percentage of disliked attributes (23.9 %). This suggests a user could dislike a dress because of this category even she likes the 'silhouette' of a dress.

**Fig. 3. (a) & (b) shows the distribution of liked and disliked attributes annotated by the users.**

**Table 2: Task 2-The number of liked and disliked attributes annotated by the users.**

<table>
<thead>
<tr>
<th>Super-category</th>
<th># Liked attribute</th>
<th>Freq.</th>
<th># Disliked attribute</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dress Style</td>
<td>2385</td>
<td>6.2 %</td>
<td>1132</td>
<td>7.7 %</td>
</tr>
<tr>
<td>Silhouette</td>
<td>8419</td>
<td>21.9 %</td>
<td>1786</td>
<td>12.2 %</td>
</tr>
<tr>
<td>Textile Pattern</td>
<td>3194</td>
<td>8.3 %</td>
<td>2031</td>
<td>13.9 %</td>
</tr>
<tr>
<td>Tex fini, manu-tech</td>
<td>5656</td>
<td>14.7 %</td>
<td>3493</td>
<td>23.9 %</td>
</tr>
<tr>
<td>None-Textile Type</td>
<td>49</td>
<td>0.1 %</td>
<td>49</td>
<td>0.33 %</td>
</tr>
<tr>
<td>Neckline Style</td>
<td>6611</td>
<td>17.2 %</td>
<td>2283</td>
<td>15.6 %</td>
</tr>
<tr>
<td>Collar Style</td>
<td>387</td>
<td>1 %</td>
<td>101</td>
<td>0.7 %</td>
</tr>
<tr>
<td>Lapel Style</td>
<td>28</td>
<td>0.1 %</td>
<td>4</td>
<td>0.02 %</td>
</tr>
<tr>
<td>Sleeve Style</td>
<td>2179</td>
<td>5.6 %</td>
<td>726</td>
<td>4.9 %</td>
</tr>
<tr>
<td>Sleeve length</td>
<td>1917</td>
<td>4.9 %</td>
<td>592</td>
<td>4.1 %</td>
</tr>
<tr>
<td>Pocket Style</td>
<td>236</td>
<td>0.6 %</td>
<td>84</td>
<td>0.5 %</td>
</tr>
<tr>
<td>Opening Type</td>
<td>691</td>
<td>1.8 %</td>
<td>453</td>
<td>3.1 %</td>
</tr>
<tr>
<td>Waistline</td>
<td>4146</td>
<td>10.7 %</td>
<td>849</td>
<td>5.8 %</td>
</tr>
<tr>
<td>Dress length</td>
<td>2499</td>
<td>6.5 %</td>
<td>982</td>
<td>6.7 %</td>
</tr>
</tbody>
</table>

**Table 3: Task 2-The number of total attributes annotated by the users for each fine-grained attribute category.**

**Fig. 3. (a) & (b) shows the distribution of liked and disliked attributes annotated by the users. The results show ‘printed’ and ‘normal waist’ are the main reason that causes some users to like a dress. However, these attributes can also be the factor that causes users dislike a dress. Because a user could like a certain type of printed pattern but dislike another type of printed pattern. To better explain a user’s fashion taste, it requires to train a model to not only understand users’ preference on a specific attribute, but also the pixel-level pattern on the area that this attribute is located. For this purpose, we asked the users to conduct task 3.**

**Task3-Human attention** Table 4 shows the number of total human attention annotated by the users for each dress length. The number of annotated attention is evenly distributed across different dress length.

**Task4-Textual explanation for task 3** Task 4 contains 11.3
counts 20 40 80 (a) Distribution of liked attributes.

the users to explain the attention that they draw for Task3. abstract) related words are high frequency words used by normal, and empire waist) and pattern (floral, geometric, frequent 1, 2, 3 grams for our dataset. Waistline (high, to calculate the frequency of words. Fig. 3.(c) shows the explanation).

task (task 4.1 for liked explanation and task 4.2 for disliked words in average per user and 5.6 words in average per sub-
task (task 4.1 for liked explanation and task 4.2 for disliked explanation).

Word count statistics: we use SGRank from Textacy [24] to calculate the frequency of words. Fig. 3.(c) shows the most frequent 1, 2, 3 grams for our dataset. Waistline (high, normal, and empire waist) and pattern (floral, geometric, abstract) related words are high frequency words used by the users to explain the attention that they draw for Task3.

Linguistic statistics: we use part-of-speech (POS) tagging from Spacy [22] to tag noun, propn, and adj of the captions annotated in Task 4. Table 5 shows the number of most frequent unique words by POS. We find the most frequent common nouns are more associated with high level description of a dress, such as neckline, pattern and waistline. In contrast, the most frequent proper nouns are more related to detailed description of a dress, such as applique, bead, and peter pan collar. This shows the linguistic diversity of our dataset.

5. Conclusion

In this work, we studied the problem of human taste in fashion product image. we introduce an explainable fashion taste dataset, Fashionpedia-Taste, with a purpose to understand fashion taste from 3 perspectives: 1) localized attribute; 2) human attention; 3) caption. The aim of this work is to enable future studies and encourage more investigation to interpretability research of fashion taste and narrow the gap between human and machine understanding of images.
Table 5: Task 4-Linguistic statistics: Number of unique words by POS.

<table>
<thead>
<tr>
<th>POS Type</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>dress, neck, length, neckline, design, shape, pattern, waist, color, sleeve, curve, waistline, fabric, skirt</td>
</tr>
<tr>
<td>Propn</td>
<td>maxi, applique, bead, kimono, peter pan, pleat, tiered halter, dolman, slit, tent, stripe, cutout, cheetah</td>
</tr>
<tr>
<td>Adj</td>
<td>elegant, beautiful, nice, cute, high, straight, perfect floral, loose, graceful, sexy, fit, attractive, charming</td>
</tr>
</tbody>
</table>

References


