Interpretable Fashion Matching using Rich Attributes

Xun Yang\textsuperscript{1}, Xiangnan He\textsuperscript{2}, Xiang Wang\textsuperscript{1}, Yunshan Ma\textsuperscript{1}, Fuli Feng\textsuperscript{1}, Meng Wang\textsuperscript{3}, Tat-Seng Chua\textsuperscript{1}

\textsuperscript{1} National University of Singapore
\textsuperscript{2} University of Science and Technology of China
\textsuperscript{3} Hefei University of Technology
Value of Fashion Industry

- 3 trillion USD, 2% of the world’s GDP in FY 2018

* Statistics are from *the State of Fashion 2018*, BOF, McKinsey & Company
Visual Fashion Computing

Our Research Focus
To determine whether a set of fashion items from different categories go well together:

- **Core:** Modeling Fashion *Compatibility*
- **Fundamental** technique to a variety of industry applications

**Fashion Matching**

**Modeling Fashion Compatibility**

- **Mix-and-Match**

- **Compatible**
- **Incompatible**

---

**Outfit Creation**
- Hsiao et al. CVPR 2018
- Han et al. MM 2017
- Feng et al. ICMR 2018

**Fashion Synthesis**
- Han et al. arxiv 2019
- Shih et al. AAAI 2018

**Fashion Recommendation**
- Song et al. MM 2017, SIGIR 2018, 2019
- Yin et al. WWW 2019
- Lin et al. WWW 2019
- Yang et al. AAAI 2019

**Complete the Look**
- [kang et al. CVPR 2019]
Traditional works on fashion compatibility primarily leverage visual appearance of items to model visual compatibility and perform matching in a latent visual space.

- Similarity/metric Learning [Veit et al. ICCV 2015; Song et al. MM 2017; Lin et al. TKDE 2019]

Weaknesses:
- Improper compatibility transferring
- Lack of interpretability

Encourage compatible items to be much closer to each other than incompatible items in a latent space.
- The **rich attributes** associated with fashion items, which describe the **semantics** of items in a **human-interpretable** way, have been largely ignored.

- Our idea: injecting **interpretability** into the **compatibility modeling** of fashion items by leveraging **rich attributes**

![Motivation Diagram]

### Rich Fashion Attributes

- **Color**: Midi-dresses, Short-sleeve, V-neckline, Summer, Floral, Viscose, Natural-white, Casual, Beach
- **Category**: Midi-dresses, Short-sleeve, V-neckline, Summer, Floral, Viscose, Natural-white, Casual, Beach
- **Style**: Tank, Multi-color, Viscose, Dry-clean, Relaxed, Stripes, Sleeveless, Summer, School
- **Shoes toe shape**: Sandals, Polyester, Red, High-heel, Open-toed, Casual, Dating
- **Occasion**: Sandals, Polyester, Red, High-heel, Open-toed, Casual, Dating
Problem Formulation
Interpretable Fashion Matching

Interpretable fashion matching

- **Input:** A corpus of fashion items with rich attributes and binary compatibility relationships \( \{X, A, Y\} \)
- **Output:**
  1. A matching function \( f: X \times X \rightarrow \mathbb{R} \), mapping a pair of items \((x_i, x_j)\) to a compatibility score
  2. A set of attribute crosses (matching patterns) that reveals which attributes in \(x_i\) and \(x_j\) dominate this matching

- **Compatibility score:** 0.8
- **Attribute crosses:**
  - [Fullbody: Category=Midi-dresses] & [Footwear: Category=Sandals]
  - [Fullbody: Style=Casual] & [Footwear: Style=Casual]
  - [Fullbody: Pattern=Floral] & [Footwear: Color=Red]

Research questions:

- How to derive such self-interpretable attribute crosses from data?
- How to learn the semantic representation of attribute crosses?
- How to unify the strengths of attribute crosses and item images?
Model Framework

Attribute-based Interpretable Compatibility (AIC)

- **Attribute-baseded Interpretable Compatibility (AIC)**
  - Tree-based Decision Rule Extraction Module
  - Attribute-based Decision Rule Embedding Module
  - Visual-Rule Joint learning Module

- **Contributions:**
  - Explicitly discover readable matching patterns from data
  - Capture the **semantics** of rich attributes
  - Self-interpretable
Model Framework

Attribute-based Interpretable Compatibility (AIC)

### Tree-based Decision Rule Extraction

- **Decision Tree**
  - A path from the root to a leaf -> a decision rule which can be seen as a higher-order attribute cross
  - Each leaf node corresponds to a decision rule, indexed by a unique rule ID

- **Boosted Tree model (Pretrained, GBDT)**
  - An ensemble of $T$ decision trees
  - Input: One-hot encoded categorical attributes of two items
  - Output: $T$ decision rules

$$r^t_{ij} : a^1_t \rightarrow a^2_t \rightarrow \cdots \rightarrow a^Z_t$$
Attribute-based Decision Rule Embedding

- **Existing solution** [Wang et al. WWW 2018]: learn the ID embedding of each rule
  - **Weak Representation**: Disregarding the semantics of each rule and cannot capture the semantic correlation between similar rules explicitly
  - **Poor Scalability**: Its parameter size is directly proportional to the size of decision rules, easily leading to overfitting when the tree number is large


© Copyright NExT++. All Right Reserved.
### Model Framework

**Attribute-based Interpretable Compatibility (AIC)**

#### Attribute-based Decision Rule Embedding

- **Existing solution** [Wang et al. WWW 2018]:
  - Learn the ID embedding of each rule (embedding lookup operation)
  - **Weak Representation**: Ignoring semantics of each rule (treat each rule independently, cannot explicitly capture the semantic correlation between similar rules)
  - **Poor Scalability**: Its parameter size is directly proportional to the size of decision rules, leading to overfitting when the tree number is large

- **Our Solution**: Learn attribute-based rule embedding by linearly modeling the attribute interactions into semantics-preserving rule embedding
  - **Lower parameter size**: Its parameter size is linear with the number of attributes
  - **Fine-grained interpretability** (e.g., second-order attribute crosses)
Model Framework

Attribute-based Interpretable Compatibility (AIC)

Visual-Rule Joint Modeling

- Learning visual embeddings of item images (pre-trained CNN)
- Reweighting decision rules with **attention network**

\[
\begin{align*}
\mathbf{w'}_{ijt} &= \mathbf{w}^T \sigma(\mathbf{W}([\mathbf{x}_i + \mathbf{r}_{ij}^t] \otimes \mathbf{x}_j, \mathbf{r}_{ij}^t]) + \mathbf{b}) \\
\mathbf{w}_{ijt} &= \frac{\exp(\mathbf{w'}_{ijt})}{\sum_{t}^{T} \exp(\mathbf{w'}_{ijt})} \\
\mathbf{r}_{ij} &= \frac{1}{T} \sum_{t=1}^{T} \mathbf{w}_{ijt} \mathbf{r}_{ij}^t
\end{align*}
\]

- Joint Modeling

\[
f(x_i, x_j, A_{ij}) = h_1^T (x_i \otimes x_j) + h_2^T r_{ij} + h_3^T ((x_i + x_j) \otimes r_{ij})
\]

\[\begin{align*}
\text{Visual} &\quad \text{Rule} &\quad \text{Visual–Rule}
\end{align*}\]
Experiment-Dataset

Attribute-based Interpretable Compatibility (AIC)

Data Source (Lookastic)

Women's Navy Check Coat, White Shift Dress, Black Leather Heeled Sandals, Black Leather Tote Bag

This smart pairing of a navy check coat and a white shift dress takes on different nuances according to how it's styled. On the footwear front, this ensemble is completed perfectly with black leather heeled sandals.

Well-matched Outfits from street style images described by multiple item attributes

We also extract more item attributes using Visenze fashion tagging tool

Experiment-Baselines

Attribute-based Interpretable Compatibility (AIC)

- **Baselines**
  - **Siamese Nets**[31] (SiaNet). It measures the visual compatibility using $\ell_2$-normalized Euclidean distance. *(Image only)*
  - **BPR-DAE**[29]. This work models the pairwise visual compatibility as the inner-product of item embeddings. *(Image only)*
  - **TransNFCM**[34]. It is a state-of-the-art fashion matching method that uses category-level complementary relationships to refine the item-item compatibility. *(Image + coarse category)*
  - **VBPR**[13]. It is a strong baseline for visually-aware user-item interaction modeling. It fuses visual information and ID embedding to enhance the item representation. *(Image + ID)*
  - **Neural Factorization Machines**[23] (NFM). It is a state-of-the-art embedding-based learning method that implicitly models higher-order feature interaction in a nonlinear way. We implement it by encoding all the item attributes and item images with embedding vectors. *(Image + attributes)*
  - **TEM**[32]. It is a state-of-the-art embedding-based learning method that combines the strength of traditional embedding-based methods and the tree-based method, which learns the ID-based embedding to represent rule. *(Image + attributes)*

- **Metrics:**
  - Mean Reciprocal Rank (MRR)
  - Hit ratio at rank $K$ (hit@$K$) ($K=5, 10$)
  - Normalized Discounted Cumulative Gain at rank (ndcg@$K$) ($K=5, 10$)

State-of-the-art methods
Experiment-Results

Attribute-based Interpretable Compatibility (AIC)

◼ Overall Comparison

Table 1: Overall Performance Comparison (%) with baseline methods. * and ** denote the statistical significance for $p_{value} < 0.05$ and $p_{value} < 0.01$, respectively, compared to the best baseline.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Lookastic-Men</th>
<th>MRR</th>
<th>hit@5</th>
<th>hit@10</th>
<th>ndcg@5</th>
<th>ndcg@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPR-DAE</td>
<td>23.35</td>
<td>30.97</td>
<td>30.90</td>
<td>23.28</td>
<td>26.17</td>
<td></td>
</tr>
<tr>
<td>Siamese</td>
<td>23.05</td>
<td>31.37</td>
<td>40.92</td>
<td>23.04</td>
<td>26.12</td>
<td></td>
</tr>
<tr>
<td>TransNFCM</td>
<td>26.14</td>
<td>34.94</td>
<td>44.27</td>
<td>26.28</td>
<td>29.30</td>
<td></td>
</tr>
<tr>
<td>YBPR</td>
<td>28.32</td>
<td>36.83</td>
<td>45.40</td>
<td>28.77</td>
<td>31.34</td>
<td></td>
</tr>
<tr>
<td>NFM</td>
<td>28.92</td>
<td>37.49</td>
<td>46.37</td>
<td>29.16</td>
<td>32.02</td>
<td></td>
</tr>
<tr>
<td>TEM</td>
<td>29.10</td>
<td>37.88</td>
<td>46.97</td>
<td>29.33</td>
<td>32.27</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>30.74**</td>
<td>39.51**</td>
<td>48.23**</td>
<td>31.06**</td>
<td>33.88**</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Lookastic-Women</th>
<th>MRR</th>
<th>hit@5</th>
<th>hit@10</th>
<th>ndcg@5</th>
<th>ndcg@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPR-DAE</td>
<td>23.69</td>
<td>32.97</td>
<td>42.25</td>
<td>24.02</td>
<td>27.02</td>
<td></td>
</tr>
<tr>
<td>Siamese</td>
<td>24.00</td>
<td>33.71</td>
<td>44.23</td>
<td>24.25</td>
<td>27.65</td>
<td></td>
</tr>
<tr>
<td>TransNFCM</td>
<td>29.88</td>
<td>41.01</td>
<td>51.08</td>
<td>30.70</td>
<td>33.96</td>
<td></td>
</tr>
<tr>
<td>YBPR</td>
<td>29.46</td>
<td>39.82</td>
<td>48.33</td>
<td>30.06</td>
<td>32.98</td>
<td></td>
</tr>
<tr>
<td>NFM</td>
<td>30.49</td>
<td>40.90</td>
<td>50.60</td>
<td>31.15</td>
<td>34.29</td>
<td></td>
</tr>
<tr>
<td>TEM</td>
<td>31.63</td>
<td>42.35</td>
<td>52.33</td>
<td>32.32</td>
<td>35.55</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>33.19**</td>
<td>43.83*</td>
<td>53.09**</td>
<td>33.94*</td>
<td>37.01**</td>
<td></td>
</tr>
</tbody>
</table>

Rel. Impro. | 5.6% | 4.3% | 2.6% | 5.8% | 4.9% | 4.9% | 3.4% | 1.4% | 5.0% | 4.0% |

Without attributes

With attributes

- Exploiting rich attributes facilitates fashion matching
- AIC achieves the best performance
- Injecting semantics into the embeddings of decision rule brings higher accuracy (AIC vs. TEM)
## Experiment-Results

**Attribute-based Interpretable Compatibility (AIC)**

### Effects of Attribute-based Decision Rule Embedding

- AIC (Attri.) consistently outperforms AIC (ID) (Attributes vs. ID)
- AIC (Attri.) performs comparable to AIC (ID), when the tree number is 5 or 10
- AIC (ID) suffers from overfitting when the tree number is large

### Table 2: Comparison (hit@5, ndcg@5, %) of the attribute-based (AIC (Attri.)) and ID-based (AIC (ID)) rule embeddings.

<table>
<thead>
<tr>
<th>TreeNum</th>
<th>Dataset Methods</th>
<th>Lookastic-Men</th>
<th>Lookastic-Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>hit@5 ndcg@5</td>
<td>hit@5 ndcg@5</td>
</tr>
<tr>
<td>T=1</td>
<td>AIC (Attri.)</td>
<td>37.16 28.92</td>
<td>42.05 32.22</td>
</tr>
<tr>
<td></td>
<td>AIC (ID)</td>
<td>35.99 27.60</td>
<td>41.05 31.27</td>
</tr>
<tr>
<td>T=5</td>
<td>AIC (Attri.)</td>
<td>39.34 30.88</td>
<td>43.66 33.80</td>
</tr>
<tr>
<td></td>
<td>AIC (ID)</td>
<td>39.05 30.59</td>
<td>43.57 33.69</td>
</tr>
<tr>
<td>T=10</td>
<td>AIC (Attri.)</td>
<td>39.51 31.06</td>
<td>43.83 33.94</td>
</tr>
<tr>
<td></td>
<td>AIC (ID)</td>
<td>39.25 30.83</td>
<td>43.46 33.78</td>
</tr>
<tr>
<td>T=50</td>
<td>AIC (Attri.)</td>
<td>39.32 30.77</td>
<td>43.81 33.97</td>
</tr>
<tr>
<td></td>
<td>AIC (ID)</td>
<td>38.85 30.33</td>
<td>42.85 33.16</td>
</tr>
<tr>
<td>T=100</td>
<td>AIC (Attri.)</td>
<td>39.45 30.90</td>
<td>43.87 34.06</td>
</tr>
<tr>
<td></td>
<td>AIC (ID)</td>
<td>37.88 29.55</td>
<td>41.99 32.38</td>
</tr>
</tbody>
</table>

### Figure 5: Comparison (MRR (%)) of the attribute-based (AIC (Attri.)) and ID-based (AIC (ID)) rule embeddings.
Experiment-Results

Attribute-based Interpretable Compatibility (AIC)

- **Effects of Attention Network**

- **Effects of Visual-Rule Joint Modeling**

  \[ f(x_i, x_j, A_{ij}) = h_1^T(x_i \otimes x_j) + h_2^T r_{ij} + h_3^T((x_i + x_j) \otimes r_{ij}) \]

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Lookastic-Men</th>
<th>Visual</th>
<th>Rule</th>
<th>Visual-Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods</td>
<td>MRR</td>
<td>hit@5</td>
<td>hit@10</td>
<td>ndcg@5</td>
</tr>
<tr>
<td>AIC (Rule only)</td>
<td>18.90</td>
<td>25.17</td>
<td>34.11</td>
<td>18.37</td>
</tr>
<tr>
<td>AIC (VRI only)</td>
<td>29.22</td>
<td>38.03</td>
<td>46.98</td>
<td>29.49</td>
</tr>
<tr>
<td>AIC (without VRI)</td>
<td>30.38</td>
<td>39.13</td>
<td>47.92</td>
<td>30.68</td>
</tr>
<tr>
<td>AIC (with VRI)</td>
<td>30.74</td>
<td>39.51</td>
<td>48.23</td>
<td>31.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Lookastic-Women</th>
<th>Visual</th>
<th>Rule</th>
<th>Visual-Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods</td>
<td>MRR</td>
<td>hit@5</td>
<td>hit@10</td>
<td>ndcg@5</td>
</tr>
<tr>
<td>AIC (Rule only)</td>
<td>23.40</td>
<td>30.97</td>
<td>39.82</td>
<td>23.30</td>
</tr>
<tr>
<td>AIC (VRI only)</td>
<td>33.12</td>
<td>43.64</td>
<td>53.28</td>
<td>33.83</td>
</tr>
<tr>
<td>AIC (without VRI)</td>
<td>32.73</td>
<td>43.19</td>
<td>52.62</td>
<td>33.43</td>
</tr>
<tr>
<td>AIC (with VRI)</td>
<td>33.18</td>
<td>43.83</td>
<td>53.09</td>
<td>33.94</td>
</tr>
</tbody>
</table>

- The attention mechanism consistently outperform max-pooling and average-pooling
- Some derived rules are invalid, thus degrading performance by simply aggregating the rule embedding
- If only using **Rule** (h2) term, AIC obtains poor accuracy
- If only using **VRI** (h3) term, AIC achieves comparable performance to (h1+h3)
- (h1+h2+h3) yields the best performance

© Copyright NExT++. All Right Reserved.
Case study on interpretability

Top-Bottom Matching

- Sophisticated knee_length top matches with low_rise bottom
- Sophisticated knee_length top doesn’t match with shorts
- Sophisticated top matches with low_rise bottom

Fullbody-Footwear Matching

- Cutout FullBody matches with Common heels Footwear
- White FullBody matches with White Footwear

AIC can discover informative matching patterns from data in a self-interpretable manner.
Conclusion

- **Pros**
  - Injecting semantics into decision rules embedding based on rich attributes
  - Modeling fashion compatibility in a self-interpretable framework

- **Cons**
  - Hard to evaluate the quality of the derived matching patterns (data-driven)

- **Future**
  - Jointly learning decision trees (attributes) and item embedding in a reinforcement learning (RL) manner (To improve generalization ability) or use fashion domain knowledge to guide the tree learning
  - Extend to personalized fashion recommendation by modeling user attributes
THANK YOU

NExT research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its IRC@SG Funding Initiative.
References

[8] Learning Type-Aware Embeddings for Fashion Compatibility. MI Vasileva et al. ECCV 2018
[10] Interpretable Partitioned Embedding for Customized Fashion Outfit Composition. Feng et al. ICMR 2018