Interpretable Fashion Matching using Rich Attributes

Xun Yang[1], Xiangnan He[2], Xiang Wang[1], Yunshan Ma[1], Fuli Feng[1], Meng Wang[3], Tat-Seng Chua[1]

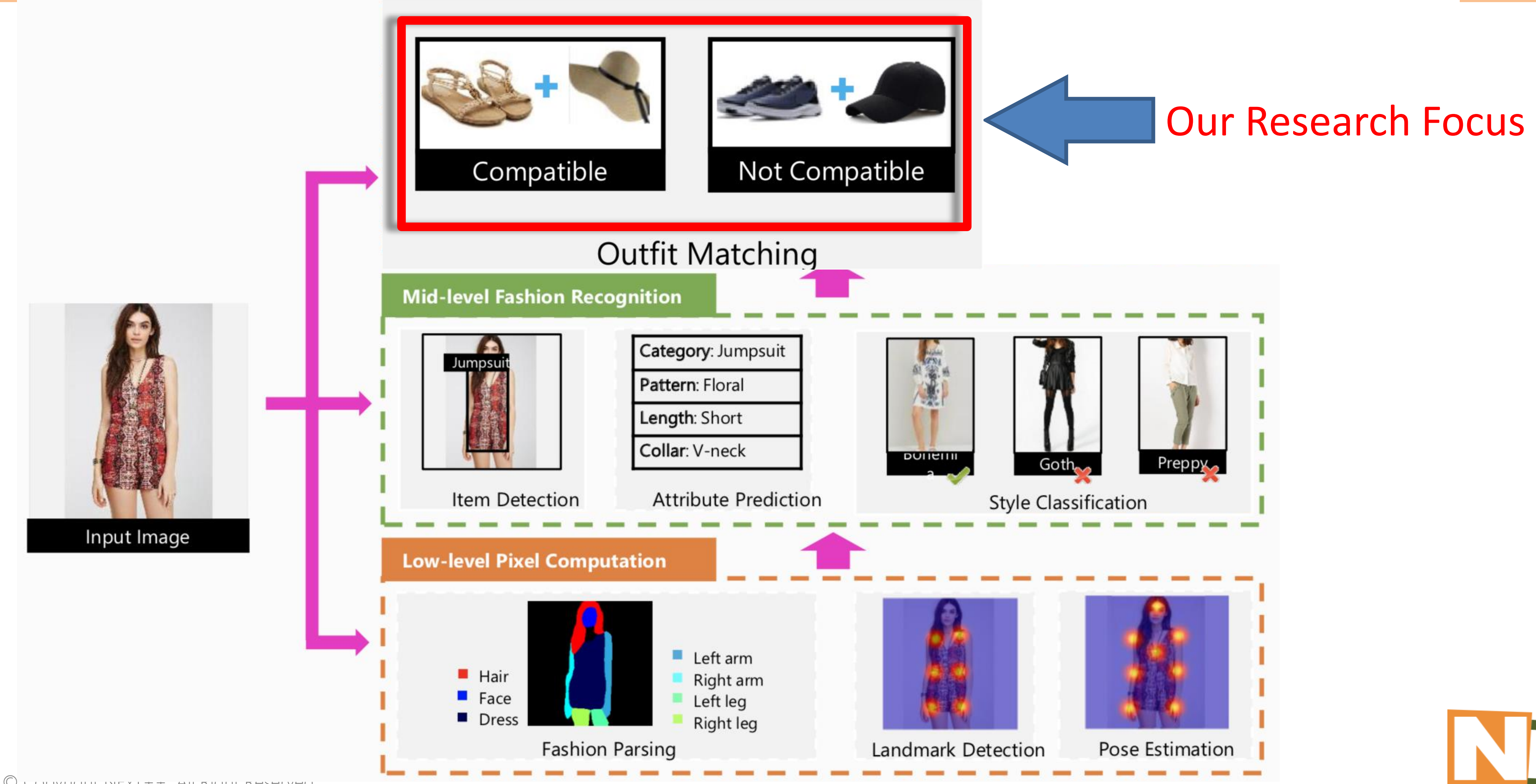
[1] National University of Singapore
[2] University of Science and Technology of China
[3] Hefei University of Technology

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

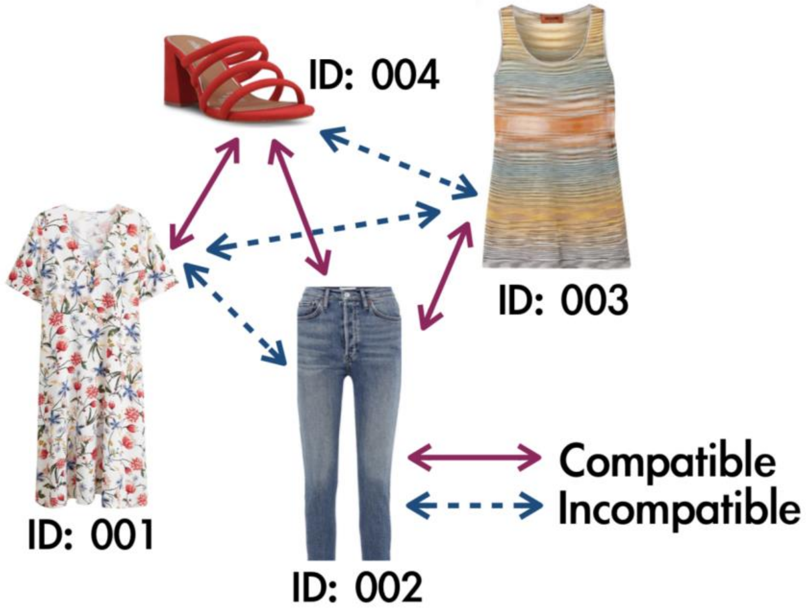
Visual Fashion Computing



Fashion Matching

Modeling Fashion Compatibility

- To determine whether a set of fashion items from different categories go well together
  - Core: <u>Modeling Fashion **Compatibility**</u>
  - **Fundamental** technique to a variety of industry applications



Mix-and-Match

Complete the Look
[kang et al. CVPR 2019]

Fashion Synthesis
[Han et al. arxiv 2019]
[Shih et al. AAAI 2018]

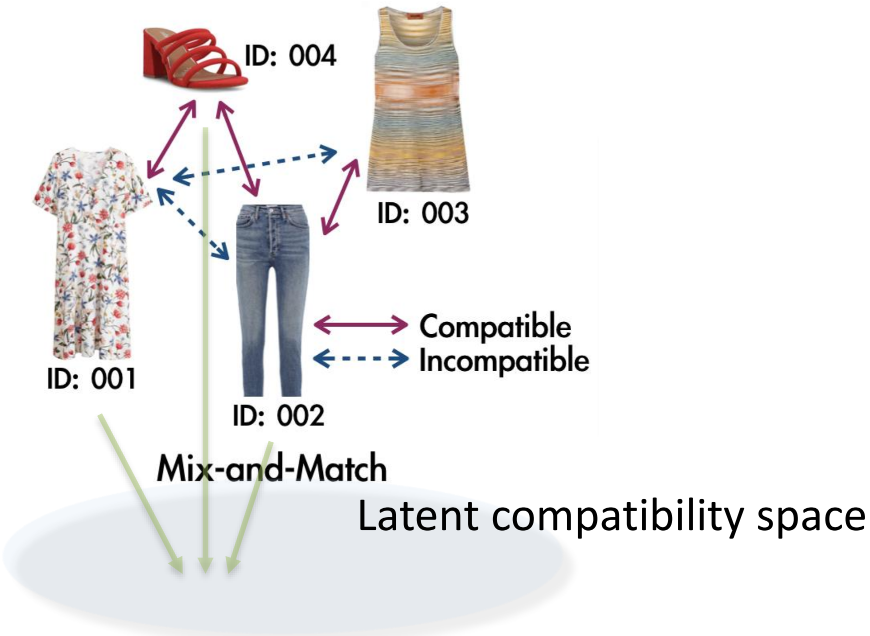
Outfit Creation
[Hsiao et. al. CVPR2018]
[Han et al. MM 2017]
[Feng et al. ICMR 2018]

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]

Existing Methods

Modeling Visual Compatibility

- 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]



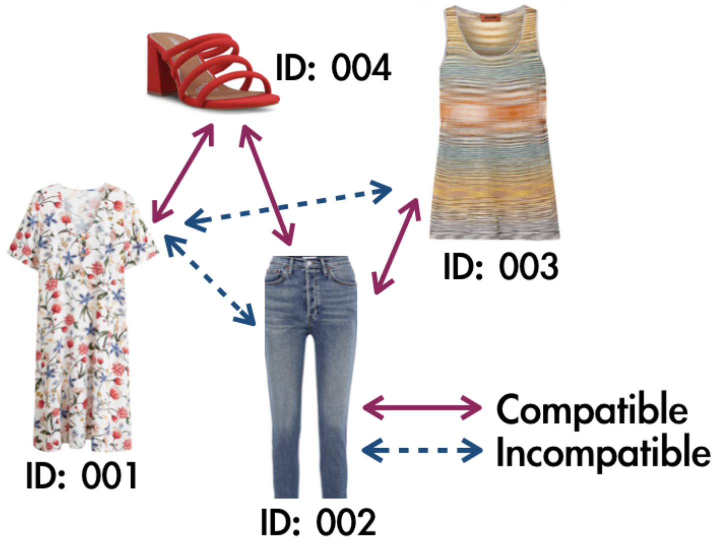
Weaknesses:

- **Improper compatibility transferring**
- **Lack of interpretability**

Encourage compatible items to be much closer to each other than incompatible items in a latent space

Motivation

- 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



Mix-and-Match

Color Category Style Shoes_toe_shape Sleeve_length Pattern Occasion Season Neckline Material Pant_rise_type

ID: 001	ID: 002	ID: 003	ID: 004
Midi-dresses	Skinny jeans	Tank	Sandals
Short-sleeve	High-rise	Multi-color	Polyester
V-neckline	Light-blue	Viscose	Red
Summer	Cotton	Dry-clean	High-heels
Floral	Denim	Relaxed	Open-toed
Viscose	Street-style	stripes	Casual
Natural-white		Sleeveless	Dating
Casual		Summer	
Beach		School	

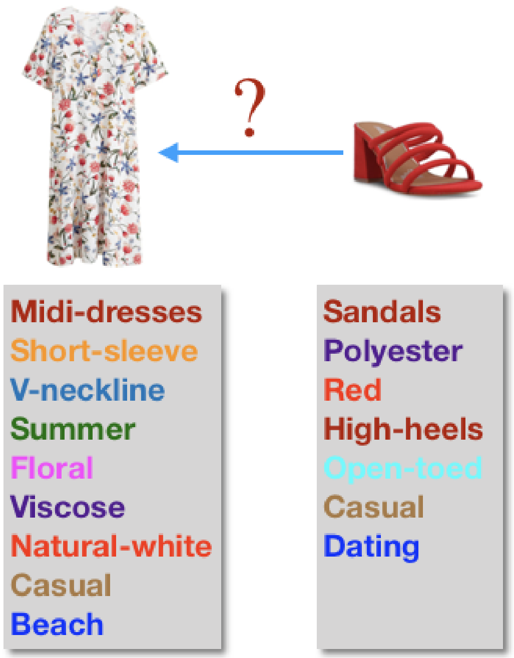
Rich Fashion Attributes

Problem Formulation

Interpretable Fashion Matching

- Interpretable fashion matching
  - Input: A corpus of fashion items with rich attributes and binary compatibility relationships $\{\mathcal{X}, \mathcal{A}, \mathcal{Y}\}$
  - Output: (1) A matching function $f: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, 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



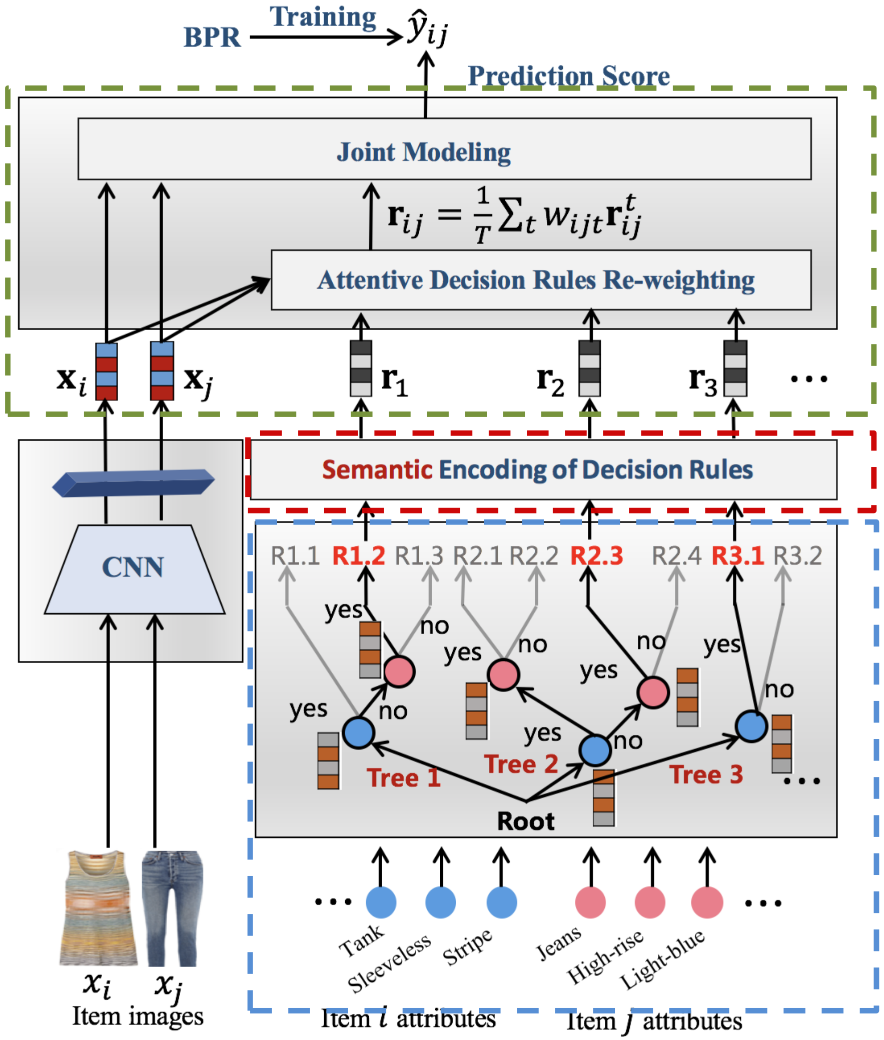
- ✓ Compatibility score: 0.8
- ✓ Attribute crosses :
  - [Fullbody: Category=Midi-dresses]&[Footwear: Category=Sandals]
  - [Fullbody: Style=Casual]&[Footwear: Style=Casual]
  - [Fullbody: Pattern=Floral]&[Footwear: Color=Red]

  ....

- 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)**

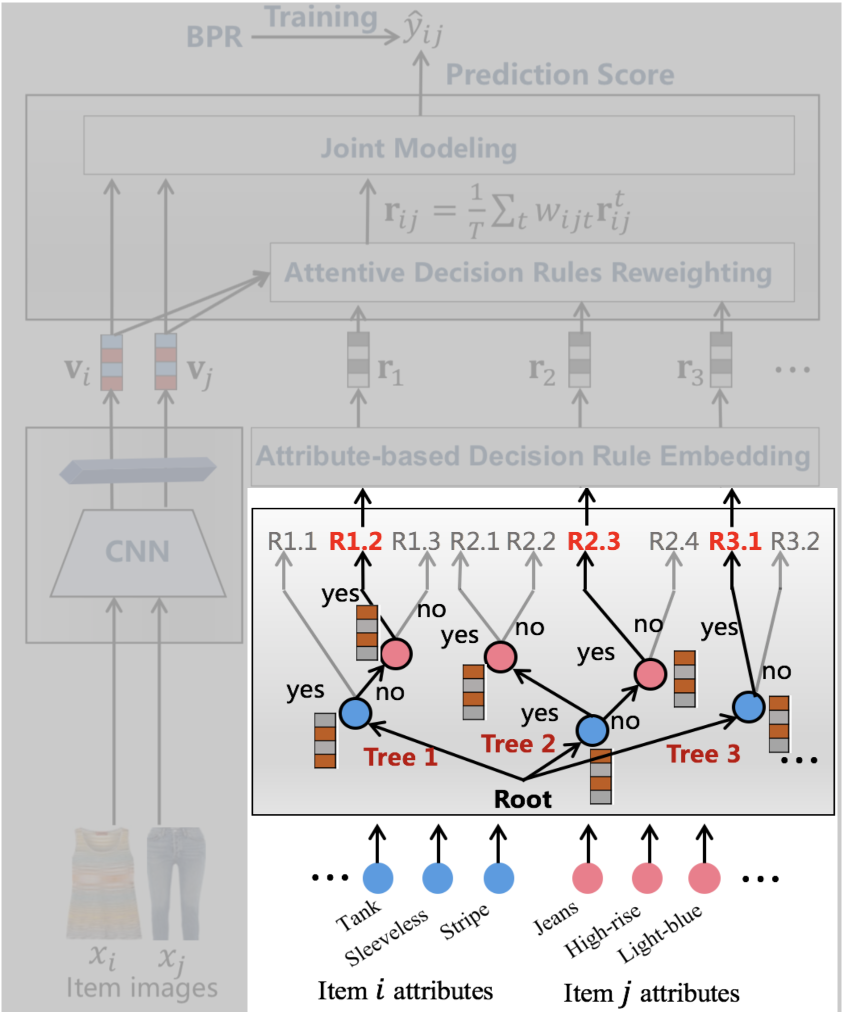
- Tree-based Decision Rule Extraction Module
- Attribute-based Decision Rule Embedding Module
- Visual-Rule Joint learning Module

- 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)

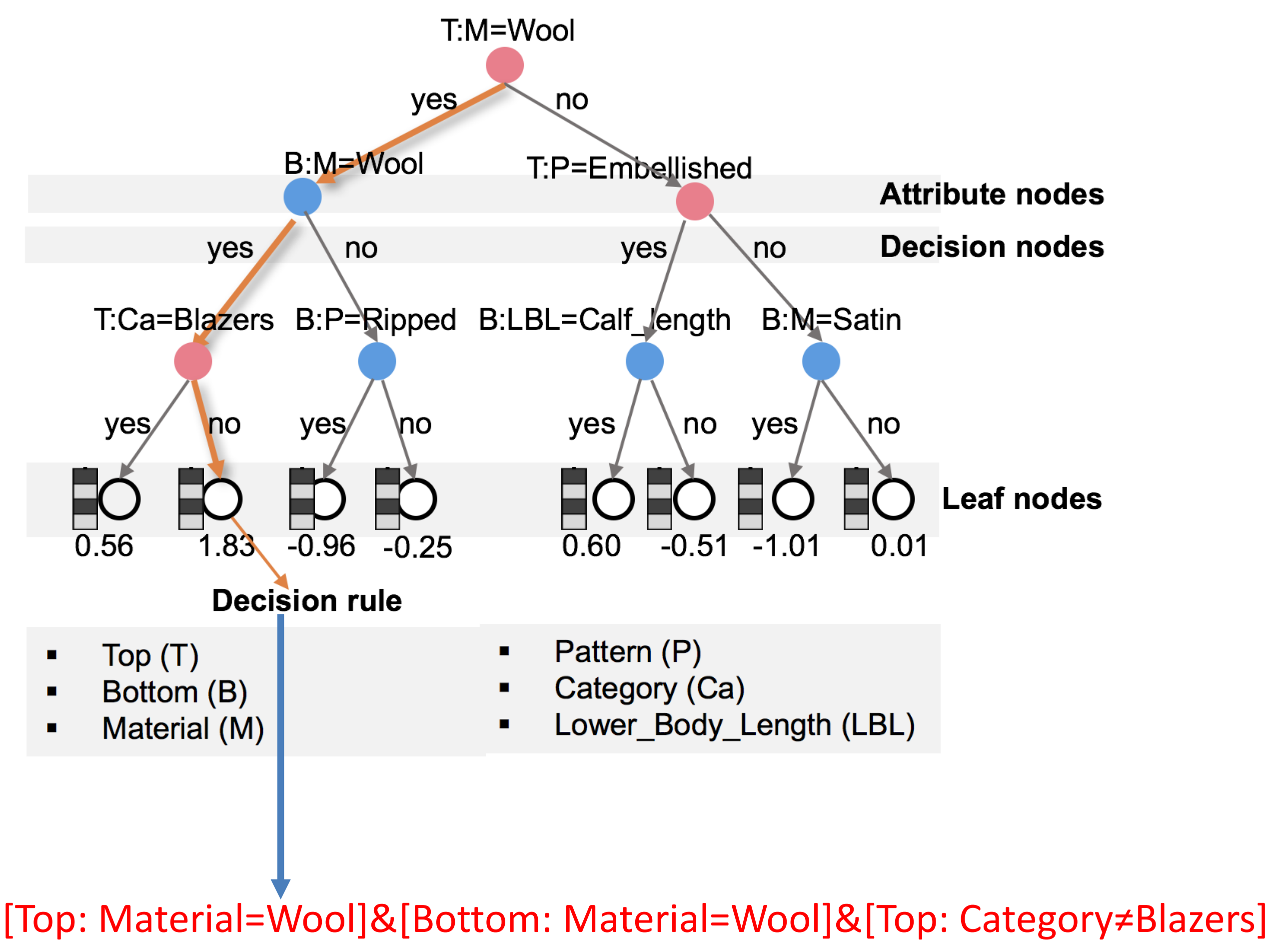
- An ensemble of T decision trees
- Input: One-hot encoded categorical attributes of two items
- Output: T decision rules

$$r_{ij}^t : a_1^t \xrightarrow{s_1^t} a_2^t \xrightarrow{s_2^t} \cdots a_Z^t \xrightarrow{s_Z^t}$$

Model Framework

Attribute-based Interpretable Compatibility (AIC)

Attribute-based Decision Rule Embedding

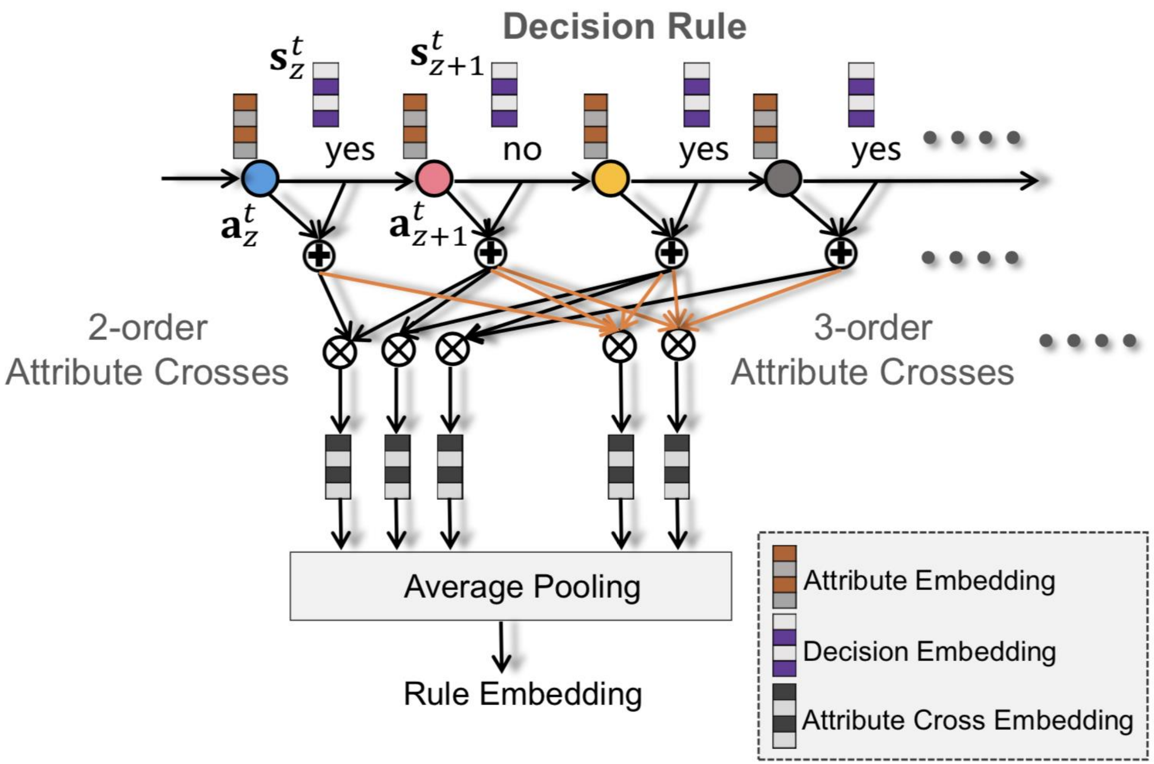


- **Existing solution** [Wang et al. WWW 2018]: learn the **ID embedding** of each rule
  - **Weak Representation**: Disregarding the **semantics** of each rule and cannot capture the semantic correlation between similar rules explicitly
  - **Poor Scalability**: Its parameter size is directly proportional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

Model Framework

Attribute-based Interpretable Compatibility (AIC)

Attribute-based Decision Rule Embedding

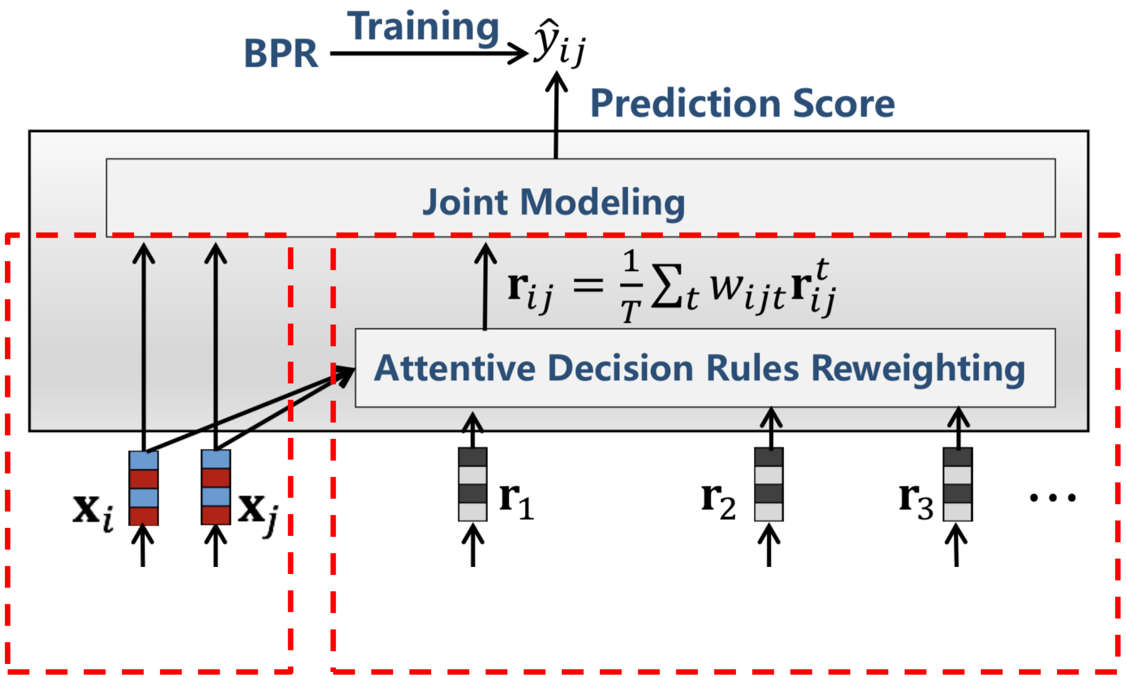


- **Existing solution**[Wang et al. WWW 2018]: learn the **ID embedding** of each rule (embedding look up operation)
  - **Weak Representation**: Ignoring **semantics** of each rule (treat each rule independently, cannot explicitly capture the semantic correlation between similar rules)
  - **Poor Scalability**: 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**

$$\begin{cases} w'_{ijt} = \mathbf{w}^T\sigma(\mathbf{W}([(\mathbf{x}_i + \mathbf{r}_{ij}^t) \otimes \mathbf{x}_j, \mathbf{r}_{ij}^t]) + \mathbf{b}) \\ w_{ijt} = \dfrac{\exp(w'_{ijt})}{\sum_t^T \exp(w'_{ijt})} \\ \mathbf{r}_{ij} = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} w_{ijt}\mathbf{r}_{ij}^t \end{cases}$$

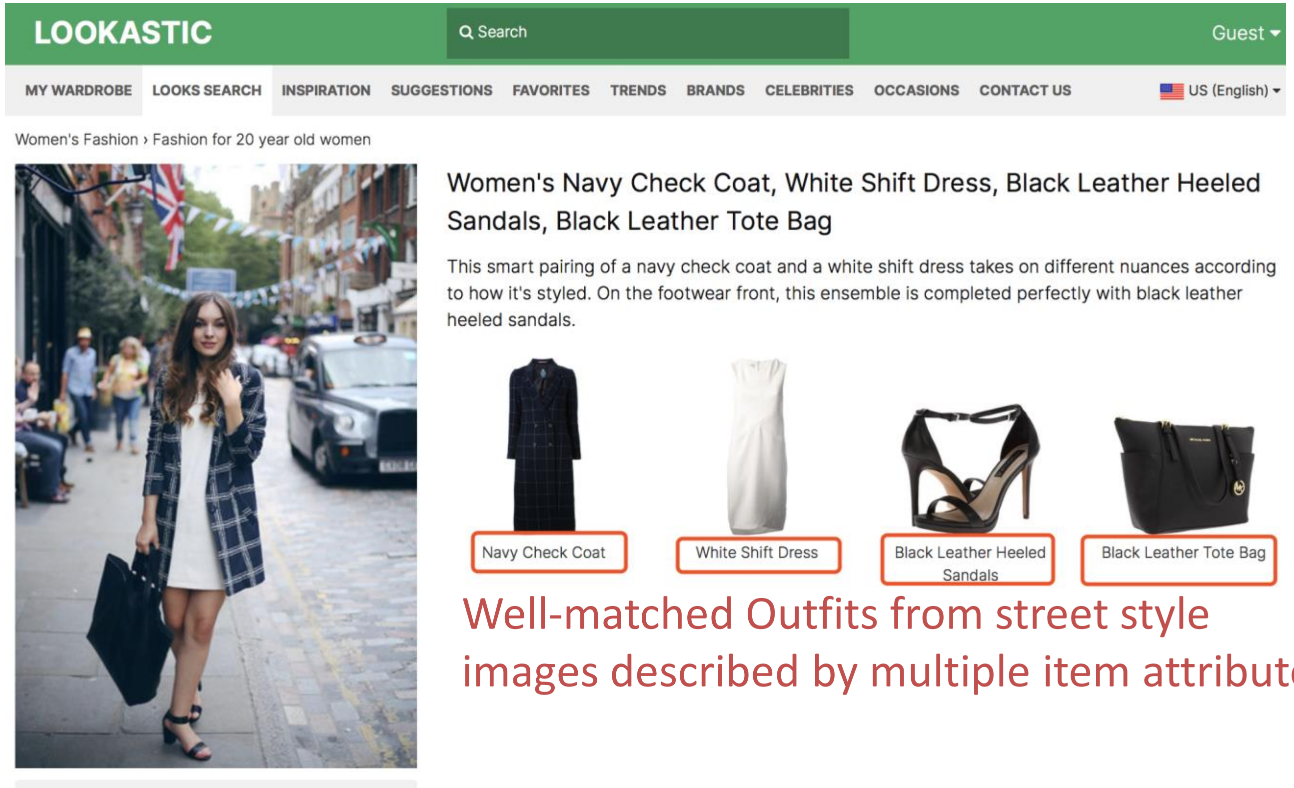
- Joint Modeling

$$f(x_i, x_j, \mathcal{A}_{ij}) = \underbrace{\mathbf{h}_1^T(\mathbf{x}_i \otimes \mathbf{x}_j)}_{Visual} + \underbrace{\mathbf{h}_2^T\mathbf{r}_{ij}}_{Rule} + \underbrace{\mathbf{h}_3^T((\mathbf{x}_i + \mathbf{x}_j) \otimes \mathbf{r}_{ij})}_{Visual-Rule}$$

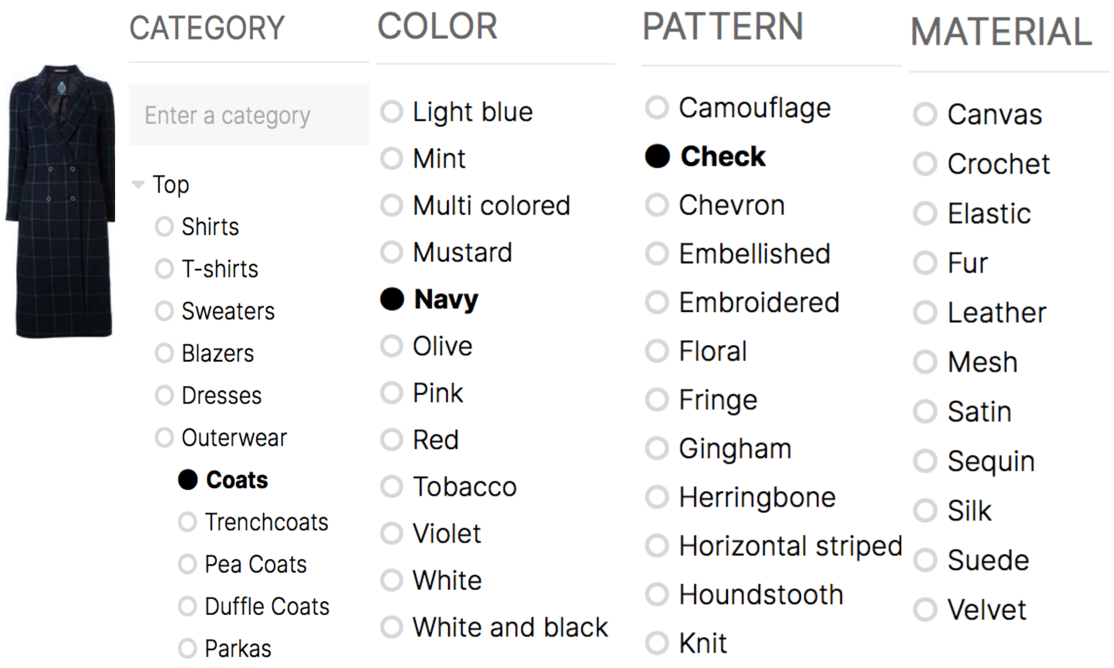
Experiment-Dataset

Attribute-based Interpretable Compatibility (AIC)

Data Source (Lookastic)



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



We also extract more item attributes using Visenze fashion tagging tool

https://www.visenze.com/automated-product-tagging

Experiment-Baselines

Attribute-based Interpretable Compatibility (AIC)

❑ Baselines

- **Siamese Nets**[31] (**SiaNet**). It measures the visual compatibility using ℓ_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**)

State-of-the-art methods

❑ 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)

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.

Dataset	Lookastic-Men				
Methods	MRR	hit@5	hit@10	ndcg@5	ndcg@10
BPR-DAE	23.35	30.97	30.90	23.28	26.17
Siamese	23.05	31.37	40.92	23.04	26.12
TransNFCM	26.14	34.94	44.27	26.28	29.30
VBPR	28.32	36.83	45.40	28.57	31.34
NFM	28.92	37.49	46.37	29.16	32.02
TEM	29.10	37.88	46.97	29.33	32.27
AIC	30.74**	39.51**	48.23**	31.06**	33.88**
Rel. Impro.	5.6%	4.3%	2.6%	5.8%	4.9%

Dataset	Lookastic-Women				
Methods	MRR	hit@5	hit@10	ndcg@5	ndcg@10
BPR-DAE	23.69	32.97	42.25	24.02	27.02
Siamese	24.00	33.71	44.23	24.25	27.65
TransNFCM	29.88	41.01	51.08	30.70	33.96
VBPR	29.46	39.32	48.33	30.06	32.98
NFM	30.49	40.90	50.60	31.15	34.29
TEM	31.63	42.35	52.33	32.32	35.55
AIC	33.19**	43.83*	53.09**	33.94*	37.01**
Rel. Impro.	4.9%	3.4%	1.4%	5.0%	4%

Without attributes

With attributes

Ours

- ❖ 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

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

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

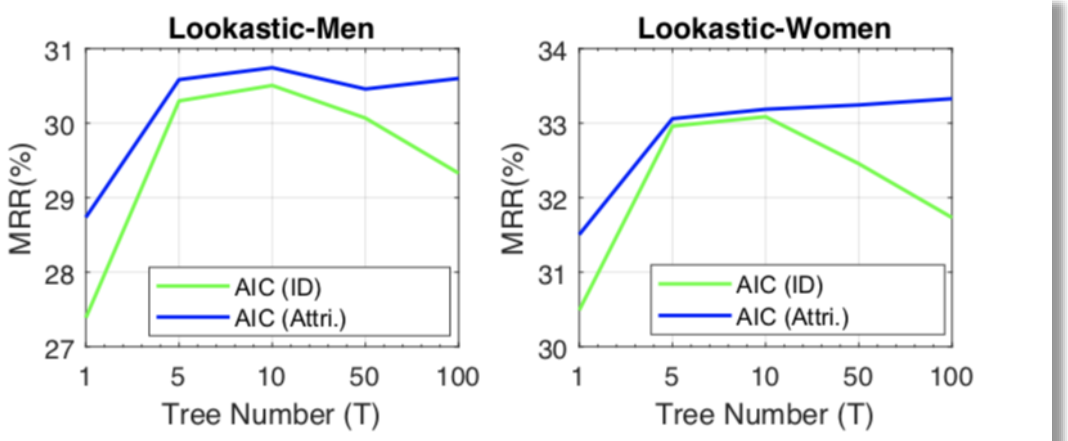


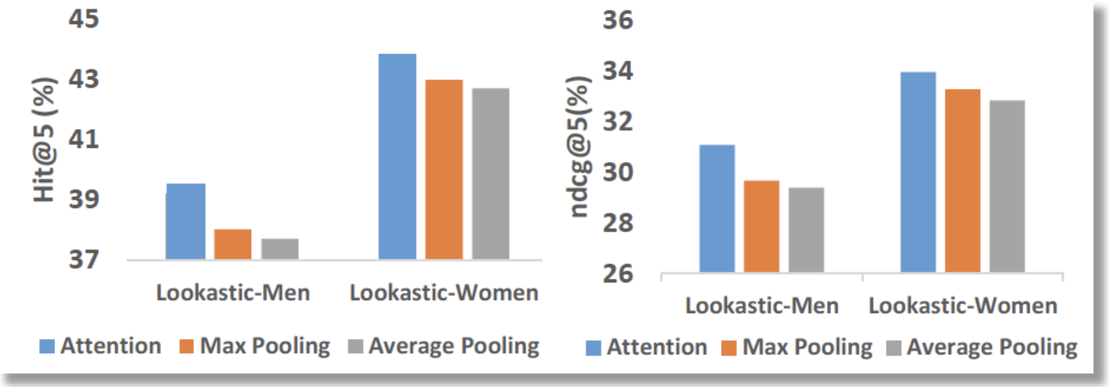
Figure 5: Comparison (MRR (%)) of the attribute-based (AIC (Attri.)) and ID-based (AIC (ID)) rule embeddings.

- ❖ 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 <u>overfitting</u> when the tree number is large

Experiment-Results

Attribute-based Interpretable Compatibility (AIC)

❑ Effects of Attention Network



- The attention mechanism consistently outperform max-pooling and average-pooling
- Some derived rules are invalid, thus degrading performance by simply aggregating the rule embedding

❑ Effects of Visual-Rule Joint Modeling

$$f(x_i, x_j, \mathcal{A}_{ij}) = \underbrace{\mathbf{h}_1^T(\mathbf{x}_i \otimes \mathbf{x}_j)}_{Visual} + \underbrace{\mathbf{h}_2^T \mathbf{r}_{ij}}_{Rule} + \boxed{\underbrace{\mathbf{h}_3^T((\mathbf{x}_i + \mathbf{x}_j) \otimes \mathbf{r}_{ij})}_{Visual-Rule}}$$

Dataset	Lookastic-Men				
Methods	MRR	hit@5	hit@10	ndcg@5	ndcg@10
AIC (Rule only)	18.90	25.17	34.11	18.37	21.25
AIC (VRI only)	29.22	38.03	46.98	29.49	32.38
AIC (without VRI)	30.38	39.13	47.92	30.68	33.52
AIC (with VRI)	30.74	39.51	48.23	31.06	33.88
Dataset	Lookastic-Women				
Methods	MRR	hit@5	hit@10	ndcg@5	ndcg@10
AIC (Rule only)	23.40	30.97	39.82	23.30	26.16
AIC (VRI only)	33.12	43.64	53.28	33.83	36.95
AIC (without VRI)	32.73	43.19	52.62	33.43	36.49
AIC (with VRI)	33.18	43.83	53.09	33.94	37.00

- 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

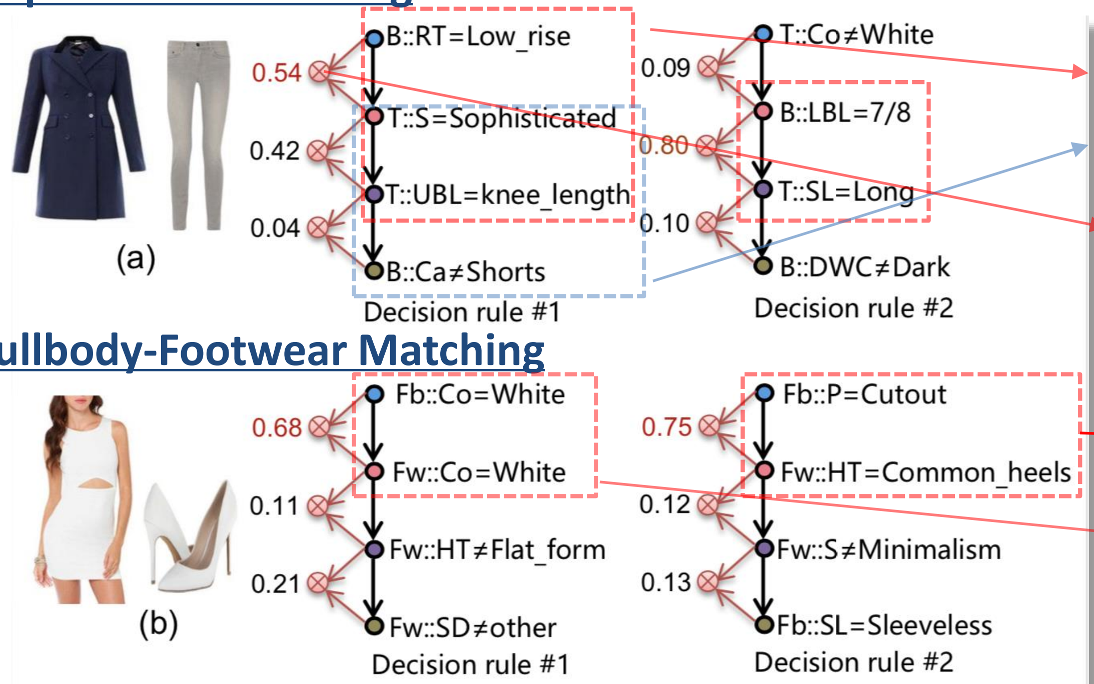
Experiment

Attribute-based Interpretable Compatibility (AIC)

❑ Case study on interpretability

Top-Bottom Matching

Fullbody-Footwear Matching



- Top (T)
- Bottom (B)
- Rise_Type(RT)
- Style (S)
- Full-body (Fb)
- Footwear (Fw)
- Heel_Type (HT)
- Shoe_Decoration (SD)
- Pattern (P)
- Category (Ca)
- Color (Co)
- Lower_Body_Length (LBL)
- Upper_Body_Length (UBL)
- Sleeve_Length (SL)
- Denim_Wash_Color (DWC)
- Shoe_Type (SD)

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

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

[1] Improving Outfit Recommendation with Co-supervision of Fashion Generation. Yujie Lin, Pengjie Ren, et al. **WWW 2019**

[2] Enhancing Fashion Recommendation with Visual Compatibility Relationship. Ruiping Yin, Kan Li, et al. **WWW 2019**

[3] Explainable Outfit Recommendation with Joint Outfit Matching and Comment Generation. Yujie Lin. IEEE TKDE 2019

[4] Complete the Look: Scene-based Complementary Product Recommendation. WC Kang, Jure Leskovec, et al. **CVPR 2019**

[5] Context-Aware Visual Compatibility Prediction. G Cucurull et al. **CVPR 2019**

[6] Creating Capsule Wardrobes from Fashion Images. Wei-Lin Hsiao, Kristen Grauman. **CVPR 2018**

[7] Learning Fashion Compatibility with Bidirectional LSTMs. Han et al. ACM MM 2017.

[8] Learning Type-Aware Embeddings for Fashion Compatibility. MI Vasileva et al. ECCV 2018

[9] Compatibility Family Learning for Item Recommendation and Generation. YS Shih et al. AAAI 2018.

[10] Interpretable Partitioned Embedding for Customized Fashion Outfit Composition. Feng et al. ICMR 2018