Prototype-guided Attribute-wise Interpretable Scheme for Clothing Matching

Xianjing Han
Shandong University
hanxianjing2018@gmail.com

Xuemeng Song*
Shandong University
sxmustc@gmail.com

Jianhua Yin
Shandong University
jhyin@sdu.edu.cn

Yinglong Wang
Shandong Computer Science Center
(National Supercomputer Center in Jinan)
wangyl@sdas.org

Liqiang Nie*
Shandong University
nieliqiang@gmail.com

ABSTRACT

Recently, as an essential part of people's daily life, clothing matching has gained increasing research attention. Most existing efforts focus on the numerical compatibility modeling between fashion items with advanced neural networks, and hence suffer from the poor interpretation, which makes them less applicable in real world applications. In fact, people prefer to know not only whether the given fashion items are compatible, but also the reasonable interpretations as well as suggestions regarding how to make the incompatible outfit harmonious. Considering that the research line of the comprehensively interpretable clothing matching is largely untapped, in this work, we propose a prototype-guided attribute-wise interpretable compatibility modeling (PAICM) scheme, which seamlessly integrates the latent compatible/incompatible prototype learning and compatibility modeling with the Bayesian personalized ranking (BPR) framework. In particular, the latent attribute interaction prototypes, learned by the non-negative matrix factorization (NMF), are treated as templates to interpret the discordant attribute and suggest the alternative item for each fashion item pair. Extensive experiments on the real-world dataset have demonstrated the effectiveness of our scheme.

CCS CONCEPTS

• **Information systems** → **Retrieval tasks and goals**; *World Wide Web*;

KEYWORDS

Fashion Analysis, Interpretable Compatibility Modeling, Non-negative Matrix Factorization.

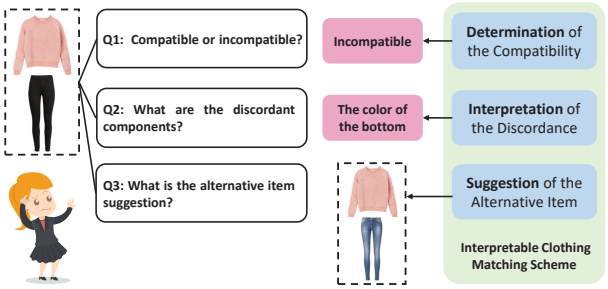


Figure 1: Illustration of the task.

ACM Reference Format:
Xianjing Han, Xuemeng Song, Jianhua Yin, Yinglong Wang, and Liqiang Nie. 2019. Prototype-guided Attribute-wise Interpretable Scheme for Clothing Matching. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19), July 21–25, 2019, Paris, France.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3331184.3331245

1 INTRODUCTION

Nowadays, clothing matching has become an indispensable part of people's daily life, since a properly coordinated outfit can improve one's appearance greatly. However, not everyone is a natural-born fashion stylist, and for those who lack the taste of aesthetics, matching clothes and making proper outfits has become their daily headache. Therefore, it is thus highly desirable to devise an automatic clothing matching scheme to aid people in outfit composition. Towards this end, three essential questions frequently faced by people in clothing matching merit our special attention. As shown in Figure 1, **Q1**: Whether the given fashion items are compatible? **Q2**: What are the discordant components that result in the incompatible matching? **Q3**: What are the alternative items to transform the incompatible pairs to compatible ones? In fact, the recent proliferation of many online fashion communities, such as IQON[1] and Chictopia[2], contributing a large number of outfits composed by fashion experts, has enabled researchers to tackle the automatic clothing matching problem. Due to their huge success in various domains, most of existing efforts employ deep learning methods to learn effective representations of fashion items, based on that they can measure the compatibility between fashion items.

* Xuemeng Song (sxmustc@gmail.com) and Liqiang Nie (nieliqiang@gmail.com) are corresponding authors.

SIGIR '19, July 21–25, 2019, Paris, France

ACM ISBN 978-1-4503-6172-9/19/07…$15.00
https://doi.org/10.1145/3331184.3331245

[1] http://www.iqon.jp/.
[2] http://www.chictopia.com/.

Nevertheless, as pure data-driven learning schemes, deep learning methods suffer from the poor interpretability given that each dimension of the learned representation cannot explicitly refer to an intuitive semantic aspect of fashion items, causing the questions **Q2** and **Q3** of requiring more result interpretations largely untapped. Notably, although a few pioneer researchers have attempted to tackle the question **Q2** by enhancing the interpretability through modeling the attribute-level (e.g. *color* and *texture*) compatibility between fashion items [7], they cannot provide the comprehensive interpretation due to the extremely limited attributes they adopt.

In this work, we aim to comprehensively tackle all the three essential problems, namely, the compatibility determination between fashion items, discordant component interpretation for incompatible outfits, and alternative item suggestion towards making compatible ones. We focus on devising a versatile attribute-wise interpretable clothing matching scheme, since attributes are the most intuitive semantic cues to characterize fashion items. However, fulfilling the task in the attribute-wise manner is non-trivial due to the following challenges. 1) As aforementioned, attribute plays a pivotal role in both characterizing fashion items and interpreting the matching results. However, most of existing benchmark datasets pertaining to clothing matching lack the attribute ground truth for fashion items. How to acquire the accurate fine-grained attribute representations for the benchmark datasets poses a primary challenge for us. 2) As the saying goes, things of one kind come together. Compatible fashion items may essentially follow certain underlying harmonious attribute interaction prototypes, while the incompatible ones would also share several unfavorable attribute compositions. For example, {*chiffon*, *pear-shaped*, *garden*, *beadings*} tends to be a harmonious attribute interaction prototype, while {*boyfriend-style*, *silk lace gauze*, *active wear*, *floral printing*} can be an incompatible one. Therefore, how to explore the latent compatible/incompatible attribute interaction prototypes and hence facilitate the discordant component interpretation is a crucial challenge. And 3) fashion items can be featured by a number of attributes, ranging from the length of trousers to the collar of the top, where each attribute further involves a set of attribute values (e.g., *long*, *short* and *mini* for the length attribute). Accordingly, the attribute interaction between fashion items can be rather complicated. How to properly model the complicated interactions among various attributes and distinguish the discordance constitutes another challenge.

To address the aforementioned challenges, we propose a prototype-guided attribute-wise interpretable compatibility modeling scheme, termed PAICM, to jointly regularize the latent prototype learning and compatibility modeling, as shown in Figure 2. Without losing the generality, here we study the problem of clothing matching between tops and bottoms. In particular, to facilitate the matching result interpretation, the scheme first extracts the semantic attribute representations for fashion items with a set of advanced neural networks, where each network is aligned to an attribute to ensure the quality of the attribute representation. Notably, to enhance the portability of PAICM, apart from our primary dataset adopted for clothing matching, we introduce an auxiliary dataset of fashion items with rich attribute annotations to pre-train the attribute classification networks. Based on the learned attribute representations, on one hand, the proposed scheme explores the latent compatible and incompatible attribute interaction prototypes using the non-negative matrix factorization (NMF) [17]. The learned prototypes are regarded as the templates to guide the discordant attribute interpretation and the alternative item suggestion. On the other hand, towards compatibility modeling, the proposed scheme seeks the latent space to accurately measure the compatibility between fashion items using the multi-layer perceptron (MLP). Ultimately, the proposed scheme seamlessly integrates the latent prototype learning and compatibility modeling with the Bayesian personalized ranking (BPR) framework [31], where the pairwise preferences between attribute prototypes and fashion items can be adaptively coupled and well exploited.

Our main contributions can be summarized in threefold:

- To the best of our knowledge, this is the first attempt to comprehensively fulfil the automatic clothing matching task by answering the three essential questions of the compatibility determination, discordant component interpretation, and alternative item suggestion.
- We propose a prototype-guided attribute-wise interpretable compatibility modeling scheme PAICM, where the latent compatible and incompatible prototype learning and compatibility modeling is jointly regularized.
- Extensive experiments have been conducted on the real-world dataset, which demonstrates the effectiveness of the proposed scheme. As a byproduct, we released the codes, and involved parameters to benefit other researchers[3].

The remainder of this paper is structured as follows. Section 2 briefly reviews the related work. In Section 3, we detail the proposed model. We present the experimental results and analyses in Section 4, followed by our concluding remarks and future work in Section 5.

2 RELATED WORK

2.1 Fashion Analyses

In recent years, the huge economic value of the fashion market has attracted many researchers' attention. Increasing efforts have been dedicated to the fashion domain, such as the fashionability prediction [19, 33], fashion trending prediction [8, 46], clothing retrieval [12, 24, 25] and compatibility modeling [9, 34, 42]. For example, McAuley et al. [28] proposed a general framework to model the human visual preference for a given pair of objects based on the Amazon real-world co-purchase dataset. In addition, Song et al. [35] investigated the problem of complementary fashion item matching with a multi-modal dataset collected from Polyvore[4]. Later, Lin et al. [20] further explored the user comments to boost the performance of fashion item recommendation, where a more comprehensive dataset ExpFashion was introduced. Although existing researches have achieved compelling success, they mainly focused on utilizing deep learning methods to represent fashion items with the blurry semantic features, resulting in their poor interpretability. To enhance the model interpretability, Feng et. al. [7] proposed a partition embedding network to learn

[3] https://anonymity2019.wixsite.com/paicm/.

[4] Polyvore has been acquired by the global fashion platform Ssense in 2018.

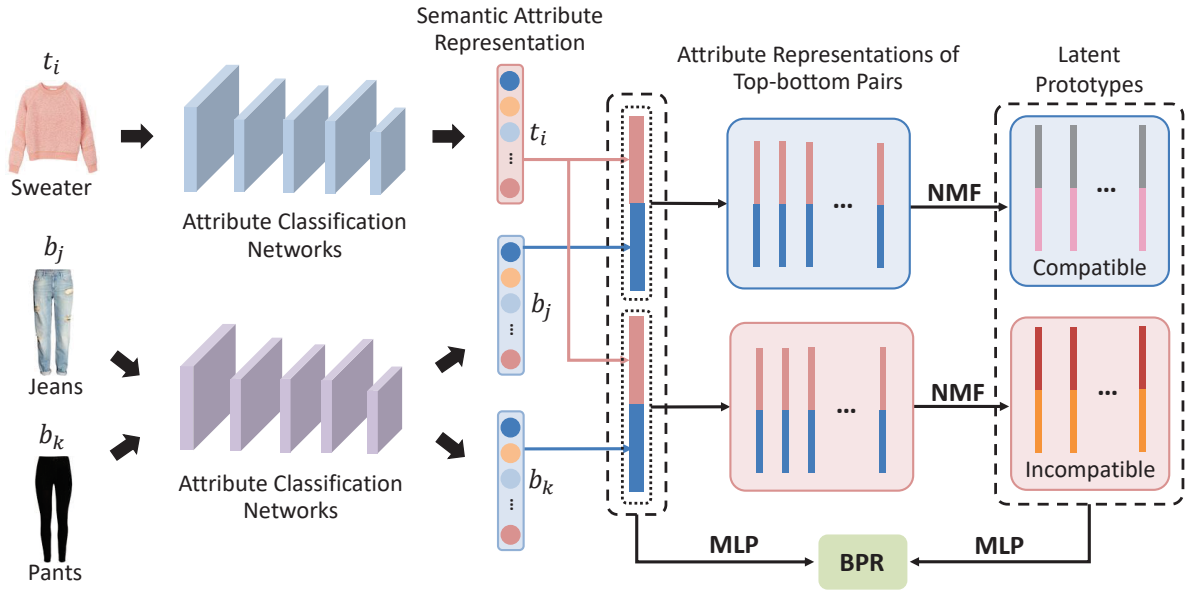


Figure 2: Illustration of the proposed scheme. We obtain the semantic attribute representations via the pre-trained attribute classification network, based on which we employ the NMF to explore the latent compatible and incompatible attribute interaction prototypes. Ultimately, we jointly regularize the latent prototype learning and compatibility modeling with the BPR framework.

the embedding of each attribute and then model the attribute-level compatibility between fashion items. Despite the promising performance it accomplished, the attributes regarding the compatibility of fashion items can be numerous yet they only adopted limited ones, making the interpretation incomprehensive. Distinguished from these studies, we aim to not only improve the interpretability of the clothing matching in a comprehensive attribute-wise manner but also facilitate the alternative item suggestion.

2.2 Matrix Factorization

As a numerical analysis method, matrix factorization (MF) is widely applied in various research areas, such as the item recommendation [11, 15, 38, 40] and information retrieval [26, 30, 37], due to its superior performance in discovering the latent features between two entities (e.g., the user and item). In order to effectively adapt to different tasks, several variants of MF have been devised, such as the singular value decomposition (SVD) [6], probabilistic matrix factorization (PMF) [29] and non-negative matrix factorization (NMF) [17], and their efficiency has been validated in various domains. For example, Sun et al. [36] proposed a SVDNet to fulfil the retrieval task of person re-identification (reID), where the SVD is employed for the optimization of the deep representation learning process. In addition, Kim et al. [14] presented a context-aware convolutional matrix factorization (ConvMF) that integrates the convolutional neural network (CNN) into the PMF in the context of document context-aware recommendation. Besides, as a useful tool for the sparse and meaningful feature extraction, NMF also drew researchers' attention. For example, Xu et al. [41] proposed a document clustering method based on the NMF with the term-document matrix. Furthermore, to forecast the fashion styles, Ziad et. al. [46] employed the NMF to discover the latent clothing styles in an unsupervised manner. Although the NMF has been successfully applied to solve tasks like text clustering [41], fashion trending prediction [46] and recommender systems [1], limited efforts have been dedicated to the complementary clothing matching, which is the major concern of our work.

3 METHODOLOGY

In this section, we first formally give the research problem formulation, and then detail the proposed PAICM.

3.1 Problem Formulation

Formally, we first declare some notations used in this work. We use bold capital letters (e.g., $\mathbf{X}$) and bold lowercase letters (e.g., $\mathbf{x}$) to denote matrices and vectors, respectively. We employ non-bold letters (e.g., x) to represent scalars and Greek letters (e.g., β) to denote the parameters. If not clarified, all vectors are in the column forms. $\|\mathbf{A}\|_F$ denotes the Frobenius norm of matrix $\mathbf{A}$.

In the real-world clothing matching scenarios, users may not only want to know whether the given fashion items are compatible or not, but also expect to get advice on how to harmonize the improper outfit. In this context, we aim to devise an attribute-wise interpretable compatibility modeling scheme to explain the underlying reasons why the given items are incompatible in the attribute-wise manner and provide the potential attribute manipulations to make compatible outfits. Assume that we have a set of tops $\mathcal{T} = \{t_1, t_2, \cdots, t_{N_t}\}$ and bottoms $\mathcal{B} = \{b_1, b_2, \cdots, b_{N_b}\}$, where N_t and N_b denote the total number of tops and bottoms,

respectively. Each item t_i (b_j) is associated with an image with a clear background, the textual description and structured category labels. In this work, we characterize each fashion item with a set of attributes (e.g., the *color* and *category*) $\mathcal{A} = \{a_q\}_{q=1}^{Q}$, where a_q is the q-th attribute and Q is the total number of attributes. Each attribute a_q is associated with a set of elements representing its possible values $\mathcal{E}_q = \{e_q^1, e_q^2, \cdots, e_q^{M_q}\}$, where e_q^i refers to the i-th element and M_q is the total number of elements regarding a_q. For simplicity, we compile all $\mathcal{E}_q$'s in order and hence derive a unified set of attribute elements $\mathcal{E} = \bigcup_{q=1}^{Q} \mathcal{E}_q = \{e_1, e_2, \cdots, e_M\}$, where $M = \sum_{q=1}^{Q} M_q$. In addition, we have a set of positive top-bottom pairs $\mathcal{S} = \{(t_{i_1}, b_{j_1}), (t_{i_2}, b_{j_2}), \cdots, (t_{i_N}, b_{j_N})\}$ composed by fashion experts, where N is the total number of positive pairs. Accordingly, for each top t_i, we can derive a set of positive bottoms $\mathcal{B}_i^+ = \{b_j \in \mathcal{B} | (t_i, b_j) \in \mathcal{S}\}$. Let s_{ij} denote the compatibility between the top t_i and bottom b_j, based on which we can distinguish whether the given fashion items are compatible or not.

3.2 Semantic Attribute Representation

As a matter of fact, the online fashion item is usually characterized by a visual image, certain user-generated textual description and structured category labels. In a sense, the visual image and structured category labels can faithfully capture the essential features of fashion items, such as the *color*, *shape* and *category*, while the user-generated textual description may be unreliable as it can be intrinsically noisy, not to mention the mendacious ones edited by crafty sellers. Therefore, similar to the existing work [46], we only exploit the reliable visual cues as well as the structured category information to model the compatibility between fashion items. Notably, existing efforts mainly adopt advanced deep neural networks to learn the effective presentations for fashion items and measure the compatibility owning to their compelling success in various research tasks. Nevertheless, as a pure data-driven learning scheme, deep neural network suffers from the poor interpretability due to the fact that each dimension of the learned representation cannot explicitly refer to the intuitive semantic aspect of fashion items. Towards this end, we aim to learn the meaningful representations for fashion items, whose dimensions directly stand for the semantic attributes and hence enhance the model interpretability.

On one hand, regarding the sophisticated visual signals, we argue that taking advantage of the well pre-trained attribute classification networks is the most natural and straightforward way to obtain the interpretable semantic representations of fashion items. As to ensure the performance of the attribute classification networks, we align each attribute a_q with a separate attribute classification network h_q. It is worth noting that as the category information also contributes an essential attribute of fashion items, here we have $Q-1$ attributes characterized by the visual cues. We feed the visual image $\mathbf{I}_i$ of the i-th top/bottom into these h_q's, and obtain the semantic attribute representations as follows,

$$\mathbf{f}_i^q = h_q(\mathbf{I}_i | \Theta_q), \quad q = 1, 2, \cdots, Q-1, \tag{1}$$

where Θ_q denotes the network parameter of h_q and $\mathbf{f}_i^q \in \mathbb{R}^{M_q}$ is the network output of h_q. The d-th entry in $\mathbf{f}_i^q$ refers to the probability that the top t_i presents the attribute element e_q^d. In particular, we denote $\mathbf{f}_i^v = [\mathbf{f}_i^1; \mathbf{f}_i^2; \cdots; \mathbf{f}_i^{Q-1}]$ as the final semantic attribute representation of the i-th top/bottom derived from the visual signals, where ";" is the cascading operation of vectors in the vertical direction.

On the other hand, the intuitive nature of the structured category information propels us to encode it directly with the one-hot representation. Let $\mathbf{f}_i^c$ stands for the one-hot semantic attribute representation derived from the category context for the i-th top/bottom. Ultimately, we concatenate the attribute representations obtained from both sources and generate the final semantic attribute representation $\mathbf{f}_i = [\mathbf{f}_i^v; \mathbf{f}_i^c]$ for the i-th item.

3.3 Latent Compatibility Space

Apparently, it is not advisable to directly measure the compatibility in the raw attribute space. Similar to [34], we assume that there is a latent compatibility space that enables us to accurately model the complicated attribute interactions and hence boost the compatibility modeling performance. In this work, we resort to the MLP, which has shown superior performance in various representation learning tasks [21–23, 39]. In particular, we add K hidden layers over the semantic attribute representation of the fashion item as follows,

$$\begin{cases} \mathbf{f}_{i0}^y = \mathbf{f}_i^y, \\ \mathbf{f}_{ik}^y = \sigma(\mathbf{W}_k^y \mathbf{f}_{i(k-1)}^y + \mathbf{b}_k^y), \quad k = 1, \cdots, K, \quad y \in \{t, b\}, \end{cases} \tag{2}$$

where $\mathbf{f}_{ik}^y$ is the k-th layer hidden representation, $\mathbf{W}_k^y$ and $\mathbf{b}_k^y$ are weight matrices and biases, respectively. t and b denote *top* and *bottom*. $\sigma : \mathbb{R} \mapsto \mathbb{R}$ is a non-linear function applied in an element-wise manner, where we choose the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$ in this work. The latent representation of the fashion item is defined as the output of the K-th layer, i.e., $\tilde{\mathbf{f}}_i^y = \mathbf{f}_{iK}^y \in \mathbb{R}^{D_l}, y \in \{t, b\}$, where D_l denotes the dimension of the latent compatibility space. Therefore, the compatibility between top t_i and bottom b_j can be measured as follows,

$$s_{ij} = (\tilde{\mathbf{f}}_i^t)^T \tilde{\mathbf{f}}_j^b. \tag{3}$$

In a sense, we can assume that the top-bottom pairs composed by fashion experts are the positive (compatible) ones. However, it may be too absolute to claim that the other fashion item pairs are negative (incompatible), since they can be the potential positive ones whose items may be paired later. In order to model the implicit relations between tops and bottoms, we adopt the BPR framework for its excellent performance on the implicit preference modeling [4, 11]. In particular, we argue that as for top t_i, bottoms in the positive set $\mathcal{B}_i^+$ are more compatible than the other bottoms. Accordingly, we construct the training set $\mathcal{D}_S := \{(i, j, k) | t_i \in \mathcal{T}, b_j \in \mathcal{B}_i^+ \wedge b_k \in \mathcal{B} \setminus \mathcal{B}_i^+\}$, where the triplet (i, j, k) indicates that top t_i goes better with bottom b_j as compared with bottom b_k. According to [31], the objective function can be written as follows,

$$\mathcal{L}_{bpr}^{item} = \sum_{(i,j,k) \in \mathcal{D}_S} -ln(\sigma(s_{ij} - s_{ik})) + \frac{\lambda}{2} \|\Omega\|_F^2, \tag{4}$$

where σ is the sigmoid function, λ is the non-negative hyperparameter to avoid the overfitting and Ω denotes the set of parameters (i.e., $\mathbf{W}_k^y$'s and $\mathbf{b}_k^y$'s).

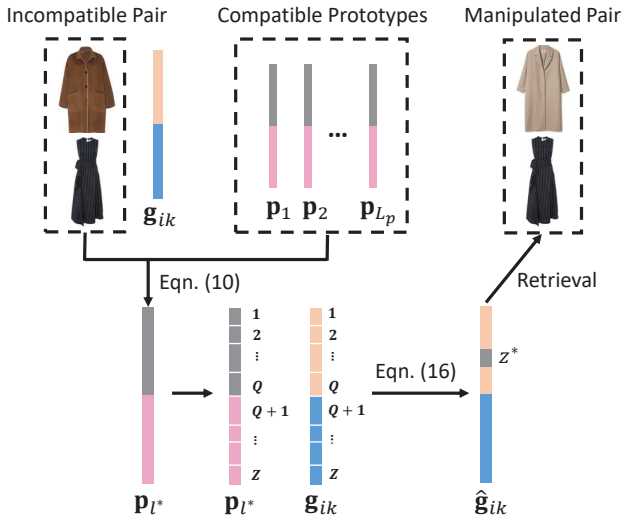


Figure 3: Workflow of the attribute manipulation.

3.4 Prototype-guided Compatibility Modeling

Intuitively, compatible fashion items can essentially follow several latent compatible attribute interaction prototypes, while the incompatible ones would share certain unfavorable prototypes. In a sense, each latent prototype can be characterized by a set of attribute elements. For example, {*jeans*, *boyfriend-style*, *ragged*, *street fashion*} tends to form a harmonious prototype, while {*office lady*, *holed*, *cartoon*, *tiered skirt*} is more likely to refer to an unfavorable one. Towards this end, we assume that there exists a set of latent compatible/incompatible attribute interaction prototypes.

Owing to its superior capability of latent factor modeling [18], we seek the latent attribute interaction prototypes under the non-negative matrix factorization (NMF). To derive the latent attribute interaction compatible prototypes, it is natural to resort to the set of positive top-bottom pairs $\mathcal{S}$. Here we define the data matrix $\mathbf{G}_p = [\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_N] \in \mathbb{R}^{2M\times N}$, where $\mathbf{g}_n = [\mathbf{f}^t_{i_n}; \mathbf{f}^b_{j_n}] \in \mathbb{R}^{2M}$ denotes the semantic attribute representation of the n-th positive top-bottom pair (t_{i_n}, b_{j_n}).

According to NMF, we aim to solve the following objective,

$$\begin{aligned} &\min_{\mathbf{P},\mathbf{H}_p} \left\|\mathbf{G}_p - \mathbf{P}\mathbf{H}_p\right\|_F^2, \\ &s.t.\ \ \mathbf{P} \geqslant 0, \mathbf{H}_p \geqslant 0, \end{aligned} \tag{5}$$

where $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_{L_p}] \in \mathbb{R}^{2M\times L_p}$ refers to the latent basis matrix, each column of which corresponds to a compatible prototype, and L_p represents the total number of the latent prototypes. $\mathbf{H}_p \in \mathbb{R}^{L_p\times N}$ corresponds to the latent representation matrix of the N top-bottom pairs regarding the basis compatibility prototypes. In particular, $\mathbf{p}_l \in \mathbb{R}^{2M}$ denotes the l-th latent compatible prototype, which can be rewritten as follows,

$$\mathbf{p}_l = \begin{bmatrix} \mathbf{p}^t_l \\ \mathbf{p}^b_l \end{bmatrix}, \tag{6}$$

where $\mathbf{p}^t_l \in \mathbb{R}^M$ and $\mathbf{p}^b_l \in \mathbb{R}^M$ can be treated as the semantic attribute representations of the prototype top and bottom for $\mathbf{p}_l$.

Algorithm 1 Prototype-guided Compatibility Modeling.

Input: $\mathcal{D}_S = \{(i, j, k)\}$, μ, υ, λ, L_p, L_u
Output: Parameters Ω in MLP, parameters $\mathbf{P}$, $\mathbf{H}_p$, $\mathbf{U}$ and $\mathbf{H}_u$ in NMF.

```
Initialize neural network parameters in MLP and NMF.
repeat
    Randomly draw (i, j, k) from D_S
    Calculate l* and r* according to Eqn. (10).
    Update Ω, P, H_p, U and H_u according to Eqn. (12).
until Converge
Identify the discordant attribute a^{z*} for the given negative top-bottom pair according to Eqn. (16)
Manipulate the discordant attribute representation and retrieve the new fashion item.
```

In the same manner, we can also derive the latent incompatible prototypes based on the set of negative top-bottom pairs (t_i, b_k)'s, where the bottom $b_k \notin \mathcal{B}^+_i$ is randomly sampled for top t_i. Let $\mathbf{G}_u \in \mathbb{R}^{2M\times N}$ be the data matrix comprising semantic attribute representations of negative top-bottom pairs and $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{L_u}] \in \mathbb{R}^{2M\times L_u}$ be the matrix of latent incompatible prototypes, where L_u is the total number of incompatible prototypes, and $\mathbf{H}_u \in \mathbb{R}^{L_u\times N}$ denotes the latent representation matrix of the N negative top-bottom pairs in the prototype space. Similarly, we represent the r-th latent incompatible prototype $\mathbf{u}_r \in \mathbb{R}^{2M}$ as follows,

$$\mathbf{u}_r = \begin{bmatrix} \mathbf{u}^t_r \\ \mathbf{u}^b_r \end{bmatrix}, \tag{7}$$

where $\mathbf{u}^t_r \in \mathbb{R}^M$ and $\mathbf{u}^b_r \in \mathbb{R}^M$ denote the semantic attribute representations of the prototype top and bottom of $\mathbf{u}_r$. Ultimately, we have the following NMF loss for the latent prototype learning,

$$\mathcal{L}_{nmf} = \left\|\mathbf{G}_p - \mathbf{P}\mathbf{H}_p\right\|_F^2 + \left\|\mathbf{G}_u - \mathbf{U}\mathbf{H}_u\right\|_F^2. \tag{8}$$

It is intuitive that the top and bottom of one compatible prototype should be more compatible than those of the incompatible ones. Therefore, we define the intrinsic compatibility for each prototype $\mathbf{p}_l$ ($\mathbf{u}_r$) as follows,

$$s^p_l = (\tilde{\mathbf{p}}^t_l)^T \tilde{\mathbf{p}}^b_l, \qquad s^u_r = (\tilde{\mathbf{u}}^t_r)^T \tilde{\mathbf{u}}^b_r, \tag{9}$$

where s^p_l and s^u_r are the intrinsic compatibility for the compatible prototype $\mathbf{p}_l$ and incompatible prototype $\mathbf{u}_r$, respectively. $\tilde{\mathbf{p}}^t_l$, $\tilde{\mathbf{p}}^b_l$, $\tilde{\mathbf{u}}^t_r$ and $\tilde{\mathbf{u}}^b_r$ are the hidden representations of $\mathbf{p}^t_l$, $\mathbf{p}^b_l$, $\mathbf{u}^t_r$ and $\mathbf{u}^b_r$, respectively, which can be acquired based on Eqn. (2).

To seamlessly integrate the latent prototype learning and compatible modeling, for each sample (i, j, k), we particularly define its most similar compatible and incompatible prototypes $\mathbf{p}_{l^*}$ and $\mathbf{u}_{r^*}$ with the Euclidean distance, whose indexes l^* and r^* can be derived as follows,

$$\begin{cases} d_p(i,j,l) = \left\| \begin{bmatrix} \mathbf{f}^t_i \\ \mathbf{f}^b_j \end{bmatrix} - \begin{bmatrix} \mathbf{p}^t_l \\ \mathbf{p}^b_l \end{bmatrix} \right\|_2, \ d_u(i,k,r) = \left\| \begin{bmatrix} \mathbf{f}^t_i \\ \mathbf{f}^b_k \end{bmatrix} - \begin{bmatrix} \mathbf{u}^t_r \\ \mathbf{u}^b_r \end{bmatrix} \right\|_2, \\ l^* = \arg\min_l d_p(i,j,l) \qquad , \quad r^* = \arg\min_r d_u(i,k,r). \end{cases} \tag{10}$$

In a sense, we expect that the intrinsic compatibility of the compatible prototype $\mathbf{p}_{l^*}$ should be higher than that of the

Table 1: Examples of attributes and the corresponding attribute elements.

Attributes	Attribute Elements
type of trousers	harem pants, straight pants
length of trousers	three quarter pants, pirate shorts
type of clothes buttons	single breasted, one button
fitness of clothes	rectangle-shaped, hourglass-shaped
length of dresses	below knee, above knee
type of dresses	A-lined dress, pouf dress
style of clothes	forest living style, boyfriend-style
texture of clothes	contrast color, hollow

incompatible one $\mathbf{u}_{r^*}$. Therefore according to the BPR, we thus have the following adaptive objective function,

$$\mathcal{L}_{bpr}^{proto} = \sum_{(i,j,k)\in\mathcal{D}_S} -ln(\sigma(s_{l^*}^{p} - s_{r^*}^{u})), \tag{11}$$

where $s_{l^*}^{p}$ and $s_{r^*}^{u}$ can be obtained with Eqn. (9). Interestingly, with $\mathcal{L}_{bpr}^{item}$ and $\mathcal{L}_{bpr}^{proto}$, the compatibility modeling between fashion items and the prototype learning can be mutually promoted. Ultimately, we obtain the final objective function as follows,

$$\mathcal{L} = \mathcal{L}_{bpr}^{item} + \mu\mathcal{L}_{bpr}^{proto} + \upsilon\mathcal{L}_{nmf}, \tag{12}$$

where μ and υ are the non-negative trade-off hyperparameters to weigh the different components of the objective function.

3.5 Interpretable Attribute Manipulation

In order to transform the incompatible fashion item pairs into the compatible ones, we first employ the L_p compatible prototypes as templates to identify the discordant attributes. In particular, for the given negative (incompatible) top-bottom pair (t_i, b_k), we particularly find the most similar compatible prototype $\mathbf{p}_{l^*}$ according to Eqn. (10). For simplicity, we divide $\mathbf{p}_{l^*}$ into Z parts as follows,

$$\mathbf{p}_{l^*} = \left[\mathbf{p}_{l^*}^{1}; \cdots; \mathbf{p}_{l^*}^{Q}; \mathbf{p}_{l^*}^{Q+1}; \cdots; \mathbf{p}_{l^*}^{Z}\right], \tag{13}$$

where $Z = 2Q$. The first Q parts refer to the attribute representations of the top in prototype $\mathbf{p}_{l^*}$, while the last Q parts correspond to that of the bottom in $\mathbf{p}_{l^*}$. In the same manner, the negative top-bottom pair (t_i, b_k) can be represented as follows,

$$\mathbf{g}_{ik} = \left[\mathbf{f}_i^t; \mathbf{f}_k^b\right] = \left[\mathbf{g}_{ik}^{1}; \cdots; \mathbf{g}_{ik}^{Q}; \mathbf{g}_{ik}^{Q+1}; \cdots; \mathbf{g}_{ik}^{Z}\right]. \tag{14}$$

Moreover, we define the attribute-wise difference between (t_i, b_k) and $\mathbf{p}_{l^*}$ as follows,

$$d_e(i, k, l^*, z) = \frac{\left\|\mathbf{g}_{ik}^{z} - \mathbf{p}_{l^*}^{z}\right\|_2}{M_z}, \tag{15}$$

where $d_e(i, k, l^*, z)$ denotes the attribute difference between (t_i, b_k) and $\mathbf{p}_{l^*}$ regarding the z-th attribute. We then identify the most discordant attribute that causes the incompatibility as follows,

$$z^* = \arg\max_{z} d_e(i, k, l^*, z). \tag{16}$$

Thereafter, to suggest the alternative item and make the compatible pair, we replace the attribute representation $\mathbf{g}_{ik}^{z^*}$ of (t_i, b_k) with $\mathbf{p}_{l^*}^{z^*}$ and hence obtain the manipulated semantic attribute representation as follows,

$$\hat{\mathbf{g}}_{ik} = \begin{cases} \left[\hat{\mathbf{f}}_i^t; \mathbf{f}_k^b\right], \text{if } z^* \le Q, \\ \left[\mathbf{f}_i^t; \hat{\mathbf{f}}_k^b\right], \text{if } z^* > Q, \end{cases} \tag{17}$$

where $\hat{\mathbf{f}}_i^t$ and $\hat{\mathbf{f}}_k^b$ are the manipulated semantic attribute representation of top t_i and bottom b_k, respectively, using which we can retrieve the new fashion items to make a compatible matching. In particular, if the discordant attribute manipulation needs to be taken on the top t_i (i.e., $z^* \le Q$), we can retrieve new tops $t_{i'}$'s by ranking the Euclidean distance d_p's between $\hat{\mathbf{f}}_i^t$ and the semantic attribute representations of training tops in the decent order. Otherwise, we can retrieve new bottoms $b_{k'}$'s by ranking d_p's between $\hat{\mathbf{f}}_k^b$ and the representations of training bottoms. The workflow of the attribute manipulation is shown in Figure 3, and the algorithm is summarized in Algorithm 1.

Table 2: Performance of attribute representation learning.

Attribute	Top	Trousers	Dress
length of upper-body clothes	0.7606	-	-
type of trousers	-	0.7233	-
part details of clothes	0.8462	0.8697	0.8181
type of clothes buttons	0.6742	-	-
length of trousers	-	0.7707	-
style of clothes	0.7698	0.7575	0.8325
fabric of clothes	0.8117	0.8738	0.8241
type of waistlines	-	0.8171	0.7798
texture of clothes	0.7668	0.8170	0.7387
graphic elements of clothes	0.7433	0.8166	0.7741
length of dresses	-	-	0.8243
design of dresses	-	-	0.8446
length of sleeves	0.7975	-	-
fitness of clothes	0.7135	-	-
type of collars	0.7839	-	-
type of dresses	-	-	0.7694
thickness of clothes	0.7668	0.8126	-
type of sleeves	0.7219	-	-
Total	0.7873	0.8280	0.8083

4 EXPERIMENT

To validate the effectiveness of the proposed model, we conducted extensive experiments on the real-world dataset FashionVC by answering the following questions:

- Does our PAICM outperform the state-of-the-art methods?
- What is the effect of NMF in the prototype-guided attribute manipulation?
- How does the proposed PAICM perform in the complementary fashion item retrieval?

In this section, we first detail the experimental settings and then illustrate the experimental results with the analyses on each above research question.

Table 3: Performance comparison among different approaches in terms of AUC.

Approach	AUC
POP	0.4206
RAND	0.5094
Bi-LSTM	0.5502
BPR-DAE	0.6026
ExIBR	0.6366
PAICM	0.7130

4.1 Experimental Settings

Dataset. To evaluate our PAICM, we adopted the public real-world dataset **FashionVC** [35] consisting of 20,726 outfits with 14,871 tops and 13,663 bottoms, composed by fashion experts. Each fashion item is associated with a visual image, relevant categories and the title description. In addition, to train the attribute classification networks and obtain the semantic attribute representations of fashion items, we utilized an auxiliary benchmark dataset of **DeepFashion** [27], comprising 33,881 fashion items, each of which is labeled by 18 attributes with 303 attribute elements. Table 1 shows several attribute examples and the corresponding attribute elements. Due to the uneven distribution of the data, we implemented the data augmentation for certain attribute classes with limited samples by multiple operations (e.g., copy, rotation and shift) with an integrated tool of Keras.

Attribute Representation Learning. Regarding the semantic attribute representation learning, we adopted the architecture similar to AlexNet [16] that consists of 5 convolutional layers followed by 3 fully-connected layers. We randomly divided the auxiliary dataset into two chunks: training set (80%) and testing set (20%), and chose the widely-used cross-entropy loss to train all the networks. We adopted the area under the ROC curve (AUC) [3] to evaluate the performance of the attribute representation learning. To gain more detailed insights, we further categorized fashion items in DeepFashion into the three groups: tops, trousers and dresses (skirts). Table 2 details the classification result of each attribute, where the last row "Total" refers to the average AUC weighted by the number of attribute elements. As can be seen, the overall performance of attribute classification with respect to AUC is satisfactory. Due to the fact that the auxiliary dataset lacks the annotations for the color attribute, for each fashion item in FashionVC, we extracted the color attribute based on the histogram calculation in the HSV space and encoded it to an one-hot vector as the color representation of the fashion item.

Parameter Tuning. We divided the positive pair set $\mathcal{S}$ into two parts: the training set $\mathcal{S}_{train}$ (80%) and testing set $\mathcal{S}_{test}$ (20%). For each positive pair (t_i, b_j), we randomly sampled three bottoms b_k's ($b_k \notin \mathcal{B}_i^+$), and each b_k corresponds to a triplet (i, j, k). We adopted the AUC [32, 45] as the evaluation metric. For optimization, we employed the stochastic gradient descent (SGD) [2]. In particular, we applied a non-negative constraint in each iteration to optimize NMF. We adopted the grid search strategy to determine the optimal values on a set of validation data temporarily split from the $\mathcal{S}_{train}$ for the regularization parameters (i.e., λ, μ and υ) among the values $\{10^r | r \in \{-4, \cdots, -1\}\}$, $[0.2, 0.4, 0.6, 0.8]$ and $[0.05, 0.1, 0.2, 0.3]$, respectively. In addition, the number of hidden units and learning rate are searched in $[128, 256, 512]$ and $[0.0001, 0.0005, 0.001]$, respectively. The proposed model is fine-tuned for 200 epochs, and the performance on the testing set is reported. We empirically found that the proposed model achieves the optimal performance with $K = 1$ hidden layer of 256 hidden units.

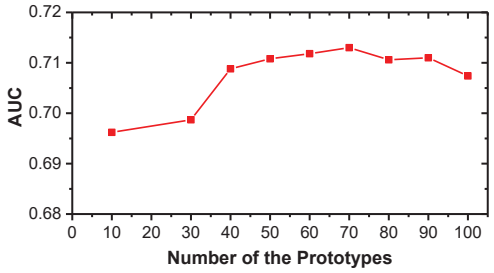


Figure 4: Performance of PAICM with respect to the number of prototypes.

4.2 On Comparison of Approaches (RQ1)

As for the compatibility modeling, we chose the following content-based baselines to evaluate the proposed model.

- **POP**: We used the "popularity" of bottom b_j to measure its compatibility with top t_i. Here the "popularity" is defined as the number of tops that has been paired with b_j in the training set.
- **RAND**: We randomly assigned the compatibility scores of s_{ij} and s_{ik} between items.
- **Bi-LSTM**: We chose the bidirectional LSTM model in [9] which explores the outfit compatibility by sequentially predicting the next item conditioned on previous ones. In our context, we adapted Bi-LSTM to deal with an outfit comprising of a top and a bottom.
- **ExIBR**: We extended the image-based recommendation (IBR) method proposed in [28] to ExIBR to handle both the visual data and the structured category label of fashion items.
- **BPR-DAE**: We selected the content-based neural scheme introduced by [35] to jointly model the coherent relation between different modalities of fashion items and the implicit preference among items via a dual autoencoder network.

To compare all the approaches fairly, we utilized both the visual image and category metadata in Bi-LSTM, ExIBR, BPR-DAE and PAICM. Table 3 shows the performance comparison among different approaches. As we can see, PAICM outperforms all the other baselines, indicating the superiority of introducing the semantic attribute representations to the compatibility modeling. One possible explanation is that the compatibility modeling task is indeed to model the complicated interactions among various attributes of fashion items, and our semantic attribute representation seems to be just task-oriented.

Moreover, as the prototype learning plays a pivotal role in our PAICM, we particularly investigate the impact of the number of the prototypes learned by the NMF on the performance of compatibility modeling. For simplicity, we adopted the same number of the compatible and incompatible prototypes, and varied that from 10 to 100 with a step of 10. Figure 4 shows the performance of our PAICM with different numbers of prototypes. We found that

Figure 5: Illustration of the compatible and incompatible prototypes. We listed several notable attributes of the prototypes according to their learned semantic representations.

the performance is relatively steady for the number of prototypes ranging from 40 to 90, where 70 is the optimal number of prototypes. This suggests that our model is not much sensitive to the number of prototypes.

To obtain the deep insights, we illustrate several learned compatible and incompatible prototypes with certain intuitive top-bottom pairs in Figure 5, where for each prototype we list the two most similar top-bottom pairs according to Eqn. (10). For clear illustration, we further give several notable attributes for each prototype based on their semantic representations. From Figure 5, we observed that the latent compatible/incompatible prototypes do share certain attribute interaction patterns. For example, "*white+black*", "*coat+dress*" and "*colledge+bow*" are the compatible attribute interactions while "*office lady+sweet*" and "*street fashion+slim fit*" are the incompatible ones. In addition, we noticed that the learned compatible prototypes are reasonable and compatible enough to be the guidance of the discordant attribute identification and the alternative item suggestion for incompatible top-bottom pairs.

4.3 On Prototype-guided Attribute Manipulation (RQ2)

To quantitatively evaluate the effects of NMF in the prototype learning, we compared NMF with K-means [10], the most commonly used unsupervised clustering method [5] that is able to group samples sharing the common characteristics together. In particular, we utilized the K-means algorithm to divide our positive top-bottom pairs into L_p clusters, and the center of each cluster is treated as the learned compatible prototype. Then according to Eqn. (10) and (16), we can find the discordant attribute and replace it with the corresponding attribute representation of the most similar compatible prototype to obtain the manipulated semantic attribute representation. As our compatibility modeling scheme PAICM is able to measure the compatibility between fashion items, here we adopted the rate of the manipulated pairs with improved compatibility as the evaluation metric. Formally, the rate is defined as $|\mathcal{M}|/|\mathcal{N}|$, where $\mathcal{N}$ denotes the set of negative top-bottom pairs determined by our PAICM model and $\mathcal{M}$ refers to the set of negative pairs, whose compatibility get improved by the attribute manipulation.

Figure 6 illustrates the performance comparison between NMF and K-means with different numbers of compatible prototypes. As can be seen, NMF consistently surpasses K-means in all configurations, demonstrating the superiority of NMF in discovering the latent prototypes. Moreover, we found that when the number of the compatible prototype is 60, we can achieve the optimal

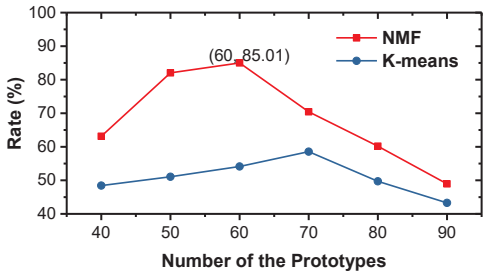


Figure 6: Performance comparison between the NMF and K-means regarding the rate of manipulated fashion item pairs with improved compatibility.

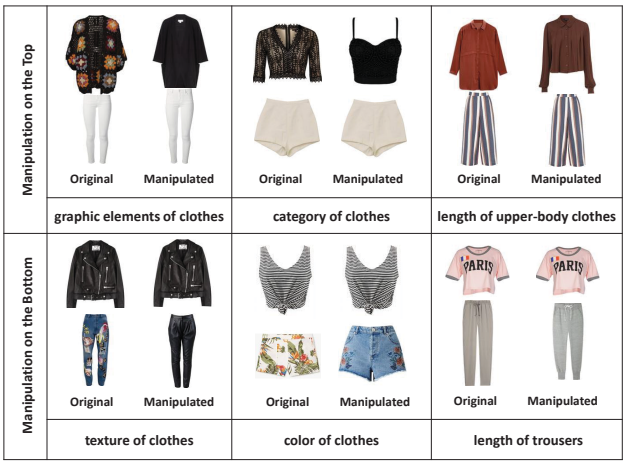


Figure 7: Illustration of the manipulated top-bottom pairs. The descriptions below the pairs are the manipulated incompatible attributes.

performance, where 85.01% of the incompatible top-bottom pairs get the compatibility improvement after the attribute manipulation. Overall, the performance is promising and validates the effectiveness of PAICM in identifying the discordant attribute and giving the reasonable alternative item suggestion. To intuitively reflect the effect of the attribute manipulation, we illustrate several examples of the manipulated top-bottom pairs in Figure 7. As we can see, the slight attribute manipulation of the incompatible top-bottom pair is able to not only improve the compatibility but also preserve the original fashion styles, which can be easily accepted by people.

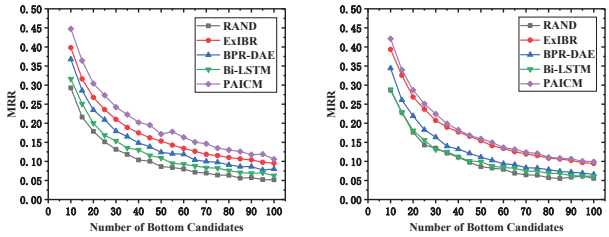
To comprehensively assess our model in attribute manipulation, apart from the above objective evaluation, we further conducted the subjective user study, where we invited 20 fashion-lovers to complete the psycho-visual test over 11 randomly selected incompatible top-bottom pairs. In particular, the attendees were asked to answer 11 independent questions by choosing the more compatible one between the original incompatible top-bottom pair and the manipulated one. All questions are presented twice to avoid the accident mistakes. The attendees taking part in the psychophysical experiment consists of 6 males and 14 females. The result of the psycho-visual test is shown in Table 4. We illustrate the maximum, minimum and average support rates of the 11 top-bottom pairs. As we can see, overall, the fashion-lovers supported the manipulated top-bottom pairs rather than the original ones, which is consistent with the above objective evaluation result.

4.4 On Fashion Item Retrieval (RQ3)

To assess the practical value of PAICM, we conducted experiments on the complementary fashion item retrieval. Considering the fact that it is time-consuming to rank all the bottoms for each top, we

Table 4: Support rate of fashion-lovers over the original and manipulated top-bottom pairs.

Support Rate	Original	Manipulated
Average Support Rate	20.68%	79.32%
Max Support Rate	46.15%	100.00%
Min Support Rate	0.00%	53.85%



(a) Observed testing tops (b) Unobserved testing tops

Figure 8: Performance of different models.

utilized the same strategy in [11] to feed each top t_i appeared in S_{test} as a query, and randomly selected T bottoms as the ranking candidates with only one positive bottom. We fed the candidates into the trained model to acquire their latent representations and calculated the compatibility score s_{ij} according to Eqn. (3), based on which we generated a ranking list of the bottoms for the given top. In this work, we focused on the average position of the positive bottom in the ranking list and thus adopted the mean reciprocal rank (MRR) metric [13, 43, 44].

In total, there are 1, 954 unique tops in the testing set. Due to the sparsity of the real-world dataset, 1, 262 (64.59%) tops never appear in $\mathcal{S}_{train}$. To comprehensively evaluate the proposed model, we divided tops in the testing set into two ground: observed testing tops and unobserved ones. As shown in Figure 8, PAICM shows superiority over all the other baselines at different numbers of bottom candidates in both scenarios, indicating the robustness and effectiveness of PAICM in complementary fashion item retrieval.

5 CONCLUSION AND FUTURE WORK

In this work, we present a prototype-guided interpretable compatibility modeling scheme, PAICM, which is capable of not only determining the outfit compatibility, but also locating the discordance of incompatible outfits as well as providing the alternative item suggestion. We employed the NMF to discover the latent compatible (incompatible) attribute interaction prototypes, which were regarded as the templates to guide the discordant attribute interpretation and alternative item suggestion. Extensive experiments have been conducted on the real-world dataset and the promising empirical results demonstrate the effectiveness of PAICM. In addition, we found that the NMF has remarkable advantages of discovering latent factors in the context of clothing matching. One limitation of our work is that currently we only manipulated the discordant attribute according to the learned prototype, but ignore the factor of users' personal preferences in clothing matching. Therefore, in the future, we plan to explore the potential of the user context in complementary clothing matching and attribute suggesting.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China, No.: 61702300, No.: 61772310, No.: 61702302, No.: 61802231, and No.: U1836216; the Project of Thousand Youth Talents 2016; the Tencent AI Lab Rhino-Bird Joint Research Program, No.:JR201805; the Future Talents Research Funds of Shandong University, No.: 2018WLJH63.

REFERENCES

[1] Jesús Bobadilla, Rodolfo Bojorque, Antonio Hernando Esteban, and Remigio Hurtado. 2018. Recommender systems clustering using bayesian non-negative matrix factorization. *IEEE Access* 6, 3549–3564.

[2] Léon Bottou. 1991. Stochastic gradient learning in neural networks. *Proceedings of Neuro-Nımes* 91, 8.

[3] Andrew P Bradley. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition* 30, 7, 1145–1159.

[4] Da Cao, Liqiang Nie, Xiangnan He, Xiaochi Wei, Shunzhi Zhu, and Tat-Seng Chua. 2017. Embedding factorization models for jointly recommending items and user generated Lists. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 585–594.

[5] Turgay Celik. 2009. Unsupervised change detection in satellite images using principal component analysis and k-means clustering. *IEEE Geoscience and Remote Sensing Letters* 6, 4, 772–776.

[6] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. 2000. A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications* 21, 4, 1253–1278.

[7] Zunlei Feng, Zhenyun Yu, Yezhou Yang, Yongcheng Jing, Junxiao Jiang, and Mingli Song. 2018. Interpretable partitioned embedding for customized multi-item fashion outfit Composition. In *Proceedings of the ACM International Conference on Multimedia Retrieval*. ACM, 143–151.

[8] Xiaoling Gu, Yongkang Wong, Pai Peng, Lidan Shou, Gang Chen, and Mohan S. Kankanhalli. 2017. Understanding fashion trends from street photos via neighbor-Constrained Embedding Learning. In *Proceedings of the ACM International Conference on Multimedia*. 190–198.

[9] Xintong Han, Zuxuan Wu, Yu-Gang Jiang, and Larry S. Davis. 2017. Learning fashion compatibility with bidirectional LSTMs. In *Proceedings of the ACM International Conference on Multimedia*. 1078–1086.

[10] John A Hartigan and Manchek A Wong. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28, 1, 100–108.

[11] Xiangnan He, Hanwang Zhang, Min Yen Kan, and Tat Seng Chua. 2016. Fast matrix factorization for online recommendation with implicit feedback. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 549–558.

[12] Diane J Hu, Rob Hall, and Josh Attenberg. 2014. Style in the long tail: Discovering unique interests with latent variable models in large scale social e-commerce. In *Proceedings of the International ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, 1640–1649.

[13] Lu Jiang, Shoou-I Yu, Deyu Meng, Yi Yang, Teruko Mitamura, and Alexander G Hauptmann. 2015. Fast and accurate content-based semantic search in 100m internet videos. In *Proceedings of the ACM International Conference on Multimedia*. ACM, 49–58.

[14] Donghyun Kim, Chanyoung Park, Jinoh Oh, Sungyoung Lee, and Hwanjo Yu. 2016. Convolutional matrix factorization for document context-aware recommendation. In *Proceedings of the ACM Conference on Recommender Systems*. ACM, 233–240.

[15] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 8, 30–37.

[16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *NIPS*. 1097–1105.

[17] Daniel D Lee and H Sebastian Seung. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 6755, 788.

[18] Xuelong Li, Guosheng Cui, and Yongsheng Dong. 2017. Graph regularized non-negative low-rank matrix factorization for image clustering. *IEEE Transactions on Cybernetics* 47, 11, 3840–3853.

[19] Yuncheng Li, Liangliang Cao, Jiang Zhu, and Jiebo Luo. 2017. Mining fashion outfit composition using an end-to-end deep learning approach on set data. *IEEE Transactions on Multimedia* 19, 8, 1946–1955.

[20] Yujie Lin, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Jun Ma, and Maarten de Rijke. 2018. Explainable fashion recommendation with joint outfit matching and comment Generation. In *arXiv preprint arXiv:1806.08977*.

[21] Meng Liu, Liqiang Nie, Xiang Wang, Qi Tian, and Baoquan Chen. 2019. Online data organizer: micro-video categorization by structure-guided multimodal dictionary learning. *IEEE Transactions on Image Processing* 28, 3, 1235–1247.

[22] Meng Liu, Xiang Wang, Liqiang Nie, Xiangnan He, Baoquan Chen, and Tat-Seng Chua. 2018. Attentive moment retrieval in videos. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 15–24.

[23] Meng Liu, Xiang Wang, Liqiang Nie, Qi Tian, Baoquan Chen, and Tat-Seng Chua. 2018. Cross-modal Moment Localization in Videos. In *Proceedings of the ACM International Conference on Multimedia*. ACM, 843–851.

[24] Si Liu, Jiashi Feng, Zheng Song, Tianzhu Zhang, Hanqing Lu, Changsheng Xu, and Shuicheng Yan. 2012. Hi, magic closet, tell me what to wear!. In *Proceedings of the ACM International Conference on Multimedia*. ACM, 619–628.

[25] Si Liu, Zheng Song, Guangcan Liu, Changsheng Xu, Hanqing Lu, and Shuicheng Yan. 2012. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*. IEEE, 3330–3337.

[26] Xin Liu, An Li, Ji-Xiang Du, Shu-Juan Peng, and Wentao Fan. 2018. Efficient cross-modal retrieval via flexible supervised collective matrix factorization hashing. *Multimedia Tools and Applications*, 1–19.

[27] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. 2016. DeepFashion: powering robust clothes recognition and retrieval with rich annotations. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1096–1104.

[28] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. 2015. Image-based recommendations on styles and substitutes. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 43–52.

[29] Andriy Mnih and Ruslan R Salakhutdinov. 2008. Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems*. 1257–1264.

[30] Dimitrios Rafailidis and Fabio Crestani. 2016. Cluster-based joint matrix factorization hashing for cross-modal retrieval. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 781–784.

[31] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the International Conference on Uncertainty in Artificial Intelligence*. AUAI Press, 452–461.

[32] Steffen Rendle and Lars Schmidt-Thieme. 2010. Pairwise interaction tensor factorization for personalized tag recommendation. In *Proceedings of the ACM International Conference on Web Search and Data Mining*. ACM, 81–90.

[33] Edgar Simo-Serra, Sanja Fidler, Francesc Moreno-Noguer, and Raquel Urtasun. 2015. Neuroaesthetics in fashion: Modeling the perception of fashionability. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*. IEEE, 869–877.

[34] Xuemeng Song, Fuli Feng, Xianjing Han, Xin Yang, Wei Liu, and Liqiang Nie. 2018. Neural compatibility modeling with attentive knowledge distillation. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 5–14.

[35] Xuemeng Song, Fuli Feng, Jinhuan Liu, Zekun Li, Liqiang Nie, and Jun Ma. 2017. NeuroStylist: neural compatibility modeling for clothing matching. In *Proceedings of the ACM International Conference on Multimedia*. ACM, 753–761.

[36] Yifan Sun, Liang Zheng, Weijian Deng, and Shengjin Wang. 2017. Svdnet for pedestrian retrieval. In *Proceedings of the IEEE International Conference on Computer Vision*. 3800–3808.

[37] Jun Tang, Ke Wang, and Ling Shao. 2016. Supervised matrix factorization hashing for cross-modal retrieval. *IEEE Transactions on Image Processing* 25, 7, 3157–3166.

[38] Thanh Tran, Kyumin Lee, Yiming Liao, and Dongwon Lee. 2018. Regularizing matrix factorization with user and item embeddings for recommendation. In *Proceedings of the ACM International Conference on Information and Knowledge Management*. ACM, 687–696.

[39] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. KGAT: Knowledge Graph Attention Network for Recommendation. In *Proceedings of the International ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

[40] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural Graph Collaborative Filtering. In *ACM SIGIR Conference on Research and Development in Information Retrieval*.

[41] Wei Xu, Xin Liu, and Yihong Gong. 2003. Document clustering based on non-negative matrix factorization. In *Proceedings of the international ACM SIGIR conference on Research and development in informaion retrieval*. ACM, 267–273.

[42] Xun Yang, Yunshan Ma, Lizi Liao, Meng Wang, and Tat-Seng Chua. 2018. TransNFCM: translation-based neural fashion compatibility modeling. In *arXiv preprint arXiv:1812.10021*.

[43] Hongzhi Yin, Hongxu Chen, Xiaoshuai Sun, Hao Wang, Yang Wang, and Quoc Viet Hung Nguyen. 2017. SPTF: A Scalable Probabilistic Tensor Factorization Model for Semantic-Aware Behavior Prediction. In *Proceedings of the IEEE International Conference on Data Mining*. 585–594.

[44] Hanwang Zhang, Zawlin Kyaw, Shih-Fu Chang, and Tat-Seng Chua. 2017. Visual Translation Embedding Network for Visual Relation Detection. In *Proceedings of the International IEEE Conference on Computer Vision and Pattern Recognition*. 3107–3115.

[45] Hanwang Zhang, Zheng-Jun Zha, Yang Yang, Shuicheng Yan, Yue Gao, and Tat-Seng Chua. 2013. Attribute-augmented semantic hierarchy: towards bridging semantic gap and intention gap in image retrieval. In *Proceedings of the ACM International Conference on Multimedia*. ACM, 33–42.

[46] Al-Halah Ziad, Stiefelhagen Rainer, and Grauman Kristen. 2017. Fashion forward: forecasting visual style in fashion. In *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, 388–397.