Personalized Fashion Recommendation via Deep Personality Learning

Dongmei Mo[1,2]
dongmei.mo@connect.polyu.hk
Xingxing Zou[1,2]
xingxingzou@aidlab.hk
Wai Keung Wong[1,2,†]
calvin.wong@polyu.edu.hk

[1] School of Fashion and Textiles
The Hong Kong Polytechnic University
Hong Kong SAR, China

[2] The Laboratory for Artificial Intelligence in Design, Hong Kong SAR, China

Abstract

Fashion personality can help individuals identify their styles' essence and make better style decisions. This paper integrates user personality with physical attributes for fashion recommendation. The proposed personality learning model (P-Net) integrates user characteristics, including personality and personal physical information (such as skin colour, hair colour, etc.), with fashion styles for a personalized recommendation. P-Net first learns outfit embeddings via a feature encoder, and the embeddings are then fed to a message-passing network to model the relations among different outfits. The personality-style learning module learns the fashion personalities of the users, and the physical compatibility is learned by exploring the relations of the feature embeddings and the physical attributes via a Transformer module. Qualitative and quantitative results on a new stylish outfit of personality (SOP) dataset indicate the superiority of P-Net compared with state-of-the-art methods and discover the potential of the combination of fashion aesthetics and psychological science. The SOP dataset is available at https://github.com/dm-mo/SOP-Stylish-Outfit-of-Personality-dataset.git.

1 Introduction

Personality understanding is effective for personalized fashion recommendation in fashion retailing. General marketing methods take questionnaire-based personality test by skilled human assessor to obtain the Myers-Briggs Personality Type Indicator (MBTI) [2] [16] types of the consumers and understand their purchasing intension and preference, and recommend outfits with different fashion styles [18]. However, the manual process is time-consuming and costly.

In the literature of personalized fashion recommendation, the existing methods are usually based on users' preference on outfit features [10], fashion patterns and designs [4], or users' historical outfit data [7]. However, they do not consider individuals' psychological types, named personality, which reflect the individuals' deeper thoughts and preference. To

†Corresponding author

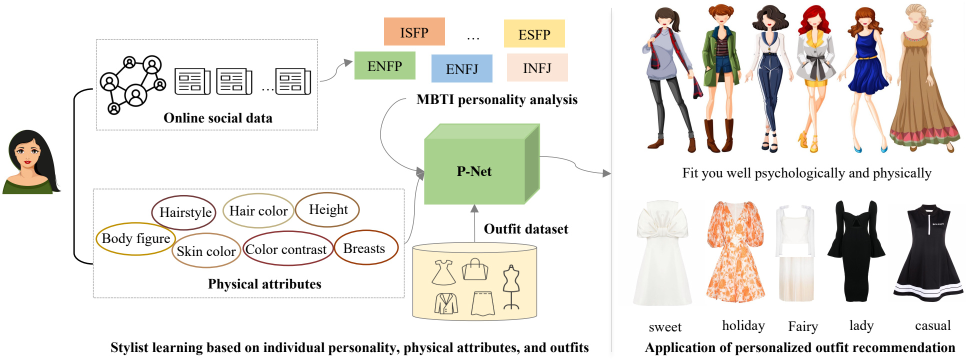

Figure 1: Concept of fashion personality learning with physical attribute support.

address this issue, we digitalize the process of consumer personality understanding and define a new task: intelligent personalized outfit recommendation via personality learning from social media data. On the one hand, we leverage the automatic personality identification model to take the place of traditional questionnaires to understand user personality traits. On the other hand, we design a deep learning model to learn the mapping between fashion styles and personality traits for personalized recommendation.

Additionally, since stylish outfit recommendation regarding individuals' physical attributes is effective for better decision making [9], we consider 7 physical aspects: body figure, skin colour, hairstyle, hair colour, height, breast size (breast) and colour contrast, to develop a P-Net for fashion personality learning. The concept of the P-Net is shown in Figure 1, in which the personality types, physical attributes and stylish outfits are jointly considered for personality-outfit mapping.

The motivation is that online social platforms have rich user behavior data that can be used to analyze users' preferences [3] and personality traits, and the preference and personality information can be well applied for personalized outfit recommendation. As shown in Figure 1, a user can be first identified with one of the 16 MBTI personality types. Then the personality information and the physical attributes, together with the user-preferred outfits, can be fed to the P-Net to learn the personality-style mapping for outfit recommendation. After training, the P-Net can be applied as an assistant to efficiently recommend outfits that satisfy the user's requirement. Figure 1 illustrates a new task: a personalized stylist to combine an individual's personality information and physical attributes for stylish outfit recommendations. The cross-domain task involves (1) personality assessment based on the user's social behaviors via integrating traditional psycholinguistic features with language model embeddings; (2) metric learning that can learn effective user embeddings and outfit embeddings for personality-style mapping in the user-style space; (3) outfit compatibility evaluation regarding individual physical attributes. This paper proposes a complete framework for dealing with this new task, and the main contributions can be summarized as:

(1) We investigate fashion personality learning and personalized recommendation by modeling online user behaviors and psychological science to understand user personality traits for recommending stylish outfits compatible with the user's physical attributes. The study has potential benefits for user understanding and personalized recommendations.

(2) We propose P-Net, an end-to-end personality-style mapping, and physical compatibility learning framework, to learn the mapping between user personality and outfit style and

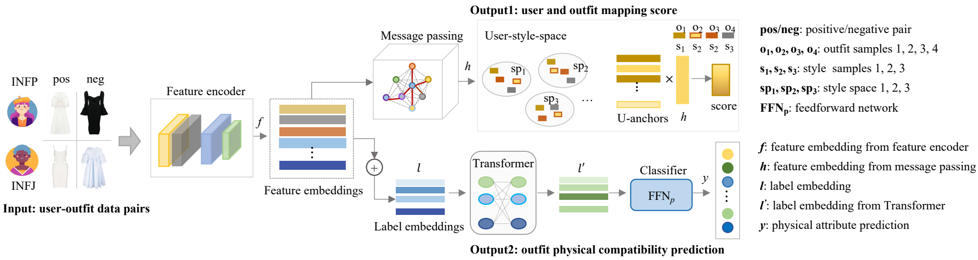

Figure 2: The proposed personalized recommendation model. P-Net contains two components: fashion personality learning (the first pipeline) and physical compatibility learning (the second pipeline). The procedure can be summarized in 3 steps: (1) The outfits corresponding to different users are input to the feature encoder to learn feature embeddings, the obtained feature embeddings are fed to the fashion personality learning pipeline and the physical compatibility learning pipeline, respectively. (2) Fashion personality learning pipeline first applies message passing network to learn the batch-based global connections of the outfits, then uses the style-aware mask strategy to learn the masked embeddings for the outfits with respect to different styles in the style space. The similarity score of the user anchors and the feature embeddings is obtained by matrix-vector multiplication. (3) Physical compatibility learning pipeline first applies Transformer encoder to model the relations between physical attributes and feature embeddings, then the FFN module is used for physical label classification.

to recommend outfits that are compatible with user physical attributes.

(3) We evaluate the performance of P-Net on the collected new stylish outfit of personality (SOP) dataset. Qualitative and quantitative results show that P-Net is superior to state-of-the-art methods, which indicates that P-Net has potential use for real online personalization recommendation systems.

Note that the related works of the proposed approach are presented in the supplementary material due to space constraints.

2 Approach

2.1 Problem Definition

Our goal is to develop a deep neural framework to discover the potential mapping between individuals' personality traits and fashion outfits, and to recommend outfits that not only best match the individuals' fashion attitude and preference, but also best fit their physical attributes. The mapping can be learned by training the framework with a large scale of data pairs: $\{u, o_{pos}, o_{neg}\}$ where u is the user embedding, $\{o_{pos}, o_{neg}\}$ represents the outfit that positively/negatively fits the user's personality. The compatibility of a user and an outfit in this paper depends on the fashion style of the outfit and the user's personality trait.

2.2 Feature Encoder of Outfit Images

The outfit can be a dress or fashion item combination. In this paper, we suppose that each outfit is composed of certain fashion items, i.e., top, bottom, shoes, and bag, then we have any outfit $o = \{x_t, x_b, x_s, x_g\}$, where $x \in R^{c \times w \times h}$, c, w, h is the channel, width and height of the fashion image, respectively. In Figure 2, the feature encoder is the Resnet18 backbone [8] pre-trained on ImageNet [5]. The output 3D tensor from the feature encoder is then transformed to 2D embedding $f \in R^{k\times d}$, where k is the feature number, and d is the feature dimensions. That is,

$$f = \Theta(o), \tag{1}$$

where $\Theta(\cdot)$ is the feature learning function. The positive and negative outfits use the same feature extraction network, and the positive and negative feature embeddings are denoted by f for simplicity.

2.3 Message Passing among Outfits

We use a message-passing network to refine the embeddings via learning global structure of the samples. The batch-based global structure of the embedding space is represented as $\mathcal{G} = (V,E)$, where V and E represent the nodes (outfits) and edges (the correlations among outfit pairs) and they are automatically learned during the training process. To fully leverage all data information in the same batch, we apply a fully connected graph (shown in Figure 2), where the nodes are initialized with feature embeddings learned from the feature extractor: $h_i^0 = f$. Similar to [17], we apply B message passing steps during the training process, and each node embedding is updated with weighted neighboring information as

$$h_i^{\beta+1} = \sum_{j=1}^{N} a_{i,j}^{\beta} W^{\beta} h_j^{\beta}, \tag{2}$$

where $a_{i,j}^{\beta}$ is the attention score of node i and j at β-th step and W^{β} is the weight matrix. Unlike most attention mechanisms that use single query and key matrices W_q, W_k, in this paper, we use T query and key matrices to transform the node features to multiple different dimensional spaces to learn a diverse set of attention scores for refining representation [17]. That is, $W_q^t \in R^{\frac{d}{T}\times d}$, $W_k^t \in R^{\frac{d}{T}\times d}$, then at the $\beta-$th step for the t-th attention score, we have

$$\alpha_{i,j}^{\beta,t} = \frac{W_q^{\beta,t} h_i^{\beta} (W_k^{\beta,t} h_j^{\beta})^T}{\sqrt{d}}, \tag{3}$$

where $W_q^{\beta,t}$ and $W_k^{\beta,t}$ are the $t-$th query and key matrices at β-step. Then, the node h_i can be updated as

$$h_i^{\beta+1} = cat(\sum_{j\in N} \alpha_{i,j}^{\beta,1} W^{\beta,1} h_j^{\beta}, \sum_{j\in N} \alpha_{i,j}^{\beta,2} W^{\beta,2} h_j^{\beta}, \ldots, \sum_{j\in N} \alpha_{i,j}^{\beta,T} W^{\beta,T} h_j^{\beta}), \tag{4}$$

where $W^{\beta,t}$ is the transformation matrix, $cat(\cdot)$ represents the concatenation function. By concatenating the attention embedding, we can have the resulting embedding $h_i^{\beta+1}$ that has the same dimension as the embedding h_i^{β}. The layer normalization is applied to the updated $h_i^{\beta+1}$, and we finally have the node embedding as

$$g(h_i^{\beta+1}) = LN(FF(f(h_i^{\beta+1})) + f(h_i^{\beta+1})), \tag{5}$$

where $f(h_i^{\beta+1}) = LN(h_i^{\beta+1} + h_i^{\beta})$ is the skip connection [8] of the attention block at β and $(\beta+1)$-th step, $LN(\cdot)$ is the layer normalization function [1].

2.4 User-style Matching Network

2.4.1 Style-aware outfit embedding

The embeddings from different fashion styles are expected to lie as far as possible. In contrast, the embeddings from the same fashion styles are expected to lie as close as possible under different style spaces for discriminant comparison. To achieve this goal, we apply a style mask strategy for the input to learn style-aware embeddings. As shown in the right side of Figure 2, the outfits from different styles (style 1, 2, 3) would have a different distance under different style spaces. For example, outfits 2 and 3 belonging to style 2 would lie far from each other under the space of style 1 or 3 while they are close to each other under the space that they belong to (style 2). For each outfit, style-aware embeddings are obtained as

$$v_i = f_{sm}(g(h_i)), \tag{6}$$

where $g(\cdot)$ denotes the massage passing function, $f_{sm}(\cdot)$ is the style mask learning function. We apply l_2 regularization on the style mask to enhance the robustness and have

$$\mathcal{L}_{mk} = \frac{\sum_{s=1}^{S} ||m_s||}{S}, \tag{7}$$

where S is the style number, m_s is the s-th style mask.

2.4.2 User-style matching

The user-style space in Figure 2 computes the user-style mapping scores with the learned user anchors and the feature embeddings. In the proposed network, each user is represented with q user anchors which are randomly initialized and automatically learned via training with the training data on the SOP dataset. The mapping score is obtained based on the matrix-vector dot product. Let a user denoted with q vectors, i.e., $u_i = \{a_1, a_2, ..., a_q\}$, the mapping scores between the user anchors and the positive feature embeddings or the compatible style mask is expected to be larger than that of the negative ones. To optimize the mapping, we have

$$s_{ij}^{u\_h} = \frac{1}{q}\sum_{i=1}^{q} a_i^T h_j, s.t. ||a_i||_2 = ||h_j||_2 = 1, \tag{8}$$

$$s_{ij}^{u\_v} = \frac{1}{q}\sum_{i=1}^{q} a_i^T v_j, s.t. ||a_i||_2 = ||v_j||_2 = 1, \tag{9}$$

$$s_{ij}^{u\_m} = \frac{1}{q}\sum_{i=1}^{q} a_i^T m_j, s.t. ||a_i||_2 = ||m_j||_2 = 1, \tag{10}$$

where $s_{ij}^{u\_h}$, $s_{ij}^{u\_v}$, $s_{ij}^{u\_m}$ are the mapping scores of the $i-$th user towards the feature embeddings of the message passing network h_j, style-aware feature embeddings v_j and the mask embeddings m_j, respectively. The log-determinant divergence (LDD) [12] is used to regularize the anchor matrix to avoid the collapse of the anchor embeddings. For an anchor matrix, $U \in R^{q \times d}$, we have

$$D_{ld}(UU^T, I) = tr(UU^T) - logdet(UU^T) - s, \tag{11}$$

where $tr(\cdot)$ is the trace of the matrix, $I \in R^{q\times q}$ is identity matrix. Suppose we have a set of outfits from the compatible styles of the $i-$th user as positive set $\mathcal{H}_i^+$, and the negative set $\mathcal{H}_i^-$ from the incompatible styles. Then, we have the training set

$$\mathcal{H} \equiv \{(i,j,k)|s_{i,j} > s_{i,k}, \forall h_j \in \mathcal{H}_i^+, h_k \in \mathcal{H}_i^-\}, \tag{12}$$

where (i,j,k) indicates that the i-th user prefers outfit j than outfit k. The Bayesian personalized ranking (BPR) [25] criteria is used to compute the loss of all users towards the corresponding outfits. For the mapping positive-negative pairs $\{s_{ij}^{u\_h}, s_{ik}^{u\_h}\}$, $\{s_{ij}^{u\_v}, s_{ik}^{u\_v}\}$, $\{s_{ij}^{u\_m}, s_{ik}^{u\_m}\}$, we have the corresponding losses:

$$\mathcal{L}_{u\_h} = \sum_{(i,j,k)\in\mathcal{H}} log(1+exp(-(s_{ij}^{u\_h} - s_{ik}^{u\_h}))), \tag{13}$$

$$\mathcal{L}_{u\_v} = \sum_{(i,j,k)\in\mathcal{H}} log(1+exp(-(s_{ij}^{u\_v} - s_{ik}^{u\_v}))), \tag{14}$$

$$\mathcal{L}_{u\_m} = \sum_{(i,j,k)\in\mathcal{H}} log(1+exp(-(s_{ij}^{u\_m} - s_{ik}^{u\_m}))). \tag{15}$$

2.5 Outfit Physical Compatibility Prediction

The physical labels can be represented with one-hot embedding: $\tilde{l} \in R^{p\times d}$, where p is the number of the physical labels. The label mask embedding strategy in [13] has been verified effective and robust for different label combinations. We apply the strategy and have the initial physical label embedding as

$$l_i = \tilde{l}_i + s_i, \tag{16}$$

where $s_i \in R^d$ indicates three states: negative (N), positive (P), and unknown (U). The state embedding is automatically learned from the learnable embedding layer.

Let the set of embeddings denoted by $Z = \{f_1, f_2, ..., f_k, l_1, l_2, ..., l_p\}$, the Transformer [19] computes the importance of each pair $\{z_i, z_j\}$ through self-attention. The attention coefficient $\tilde{\alpha}_{i,j}$ between z_i and z_j is computed by $\tilde{\alpha}_{i,j} = softmax((\tilde{W}^q z_i)^T(\tilde{W}^k z_j)/\sqrt{d})$, where $\tilde{W}^q$ and $\tilde{W}^k$ are the query and key matrices, respectively. Then we update the $i-$th embedding with weighted sum of M related embeddings as $\bar{z}_i = \sum_{j=1}^{M} \tilde{\alpha}_{i,j}\tilde{W}^v z_j$, where $\tilde{W}^v$ is the value matrix. The *ReLu* activation function is used to obtain the activated embedding as $\widetilde{z}_i = ReLU(\bar{z}_i W^r + b_1)W^o + b_2$, where W^r, W^o are transformation matrices while b_1, b_2 are bias vectors. The Transformer is repeated with L layers ($L = 3$). We denote the output of the Transformer as $Z' = \{f_1', f_2', ..., f_k', l_1', l_2', ..., l_p'\}$ and feed the updated label embeddings l_p' to the feedforward network (FFN_p) for physical label prediction. The FFN_p comprises a single linear layer followed by an activation function:

$$y_i = \mathrm{FFN}_p(l_i') = \sigma(w_i \cdot l_i' + b_i), \tag{17}$$

where w_i is the weight variable, b_i is the bias and $\sigma(\cdot)$ is the sigmoid function. The binary cross entropy loss function is used to optimize the physical label prediction and we have

$$\mathcal{L}_p = \sum_{i=1}^{N} \mathcal{E}_{p(y_k)}\{BCE(y_i', y_i)|y_k\}, \tag{18}$$

where y_k denotes the known labels, y_i and y_i' are the predicted label and the ground truth, $BCE(\cdot)$ is the binary cross entropy loss function, $\mathcal{E}_{p(y_k)}(\cdot|y_k)$ denotes the expectation regarding the probability distribution of y_k.

Combining the Eqs. (7), (11), (13), (14), (15) and (18), we have the overall loss of the proposed model as

$$\mathcal{L} = \mathcal{L}_{u\_h} + \mathcal{L}_{u\_v} + \mathcal{L}_{u\_m} + \mathcal{L}_{mk} + \gamma D_{ld} + \lambda \mathcal{L}_p, \tag{19}$$

where γ, λ are the parameters to balance the loss terms.

3 Experiment

This paper focuses on solving two tasks: one is to recommend outfits that attract a user most from the style aspect, and the other is to evaluate if the outfit can fit the user from the physical aspect. The two tasks can be considered as metric learning and multi-label classification problems. This section evaluates the performance of P-Net for solving the two tasks on the SOP dataset. More details about the SOP dataset and the experimental settings and additional results are shown in the supplementary material.

3.1 Dataset

3.1.1 Stylish Outfit of Personality (SOP) dataset.

We newly establish a large-scale dataset to facilitate the personality learning task. It contains 969 users annotated with 16 MBTI personality traits and corresponding personality-outfit mapping pairs where the outfit data are from the Outfit for You (O4U) dataset [11]. The user personality trait is obtained by fedding tweet data to a psycholinguistic learning model [15] for personality prediction. The style prediction model [6] pre-trained on the DeepFashion [14] dataset, is used for style recognition on the O4U dataset. Each outfit is labeled with a single fashion style category. Eight fashion experts from the background of fashion design and business map the 16 MBTI personality traits to different fashion styles. The mapping between personality traits and fashion styles is denoted as binary identification, where the positive indicates the styles match the user's personality while those unmatched are negative. Finally, the mapping between user personality and fashion styles can be reflected via user-outfit data pairs, i.e. [User idx, Pos/Neg outfit id]. The data pairs along with outfit images are input to the proposed P-net and the compared methods for model training.

The outfits on the O4U dataset are divided into two parts: the training part and the testing part, with a proportion of $8:2$ to avoid overlapping. We create user-outfit data pairs for SOP dataset by randomly selecting positive and negative outfits for each user. Let η denote the number of user-outfit pairs for each user, and μ denote the actual amount of outfits of a style. We set $\eta = 200$ (or $\eta = \mu$ if $\eta > \mu$) and randomly choose η outfits for each user from the training and testing parts to construct the training and testing sets, respectively. The amount of positive/negative user-outfit pairs on the training set is $161,014$ while that on the testing set is $151,911$. The training set is further divided to train, val and test subsets for model training while the testing set is only used in the testing stage. Table 1 lists the details of the SOP dataset, where Testing100 denotes the testing set with $\eta = 100$. Please note that the SOP dataset can only be partially open as we need to protect the users' privacy, while the user-outfit data pairs can be open for fashion personality learning as they are not directly related to any user's private information.

Data	Train	Val	Test	Training	Testing100	Testing
Positive	112,582	15,731	32,701	161,014	79,412	151,911
Negative	112,582	15,731	32,701	161,014	79,412	151,911

Table 1: Positive/negative user-outfit pairs on SOP dataset.

Method	Testing100 ACC	Testing100 NDCG	Testing ACC	Testing NDCG
Resnet18	0.8360	0.8357	0.8432	0.8430
CSN	0.6175	0.5660	0.6047	0.5532
T-Aware	0.6763	0.5614	0.6820	0.5569
SCE-Net	0.5389	0.5809	0.5383	0.5826
MCN	0.5197	0.6085	0.5184	0.6128
TDRG	0.5157	0.6110	0.5361	0.6069
LAPE	0.6348	0.7425	0.6395	0.7464
Ours	**0.9242**	**0.9380**	**0.9309**	**0.9434**

Table 2: Comparison of different methods.

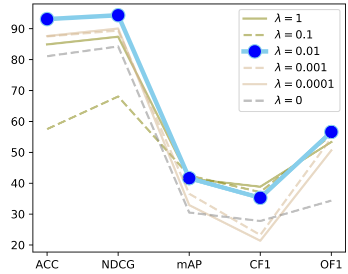

Figure 3: Performance vs. λ.

3.2 Experimental Results

In the experiments, the inputs of the compared methods and the P-Net are user-outfit data pairs. We present qualitative and quantitative analysis and explore key experimental findings from the following aspects:

3.2.1 Does P-Net outperform state-of-the-art methods for personality-style mapping?

As shown in Table 2, P-Net obtains the highest ACC and NDCG and presents superiority in learning comprehensive metrics for comparing outfits from compatible/incompatible fashion styles. The potential reason for the high performance of P-Net is that it applies a message-passing network during the training process to capture the potential relationship among different data pairs and the learned relationship information is then incorporated in the feature embedding for mapping comparison in the user-style space. Resnet18 obtains competitive ACC and NDCG compared with the comparison methods due to the optimization of the binary cross entropy loss for the metric learning problem while methods like CSN, type-aware and SCE-Net based on margin loss cannot address the problem satisfactorily.

Method	Testing100 mAP	CP	CR	CF1	OP	OR	OF1	Testing mAP	CP	CR	CF1	OP	OR	OF1
Resnet18	31.39	12.12	20.00	15.09	58.63	38.59	46.54	31.40	12.44	20.00	15.34	59.01	38.79	46.81
CSN	36.94	34.22	29.66	31.78	57.60	**51.45**	54.35	37.76	39.04	30.95	34.53	57.86	48.82	52.96
T-Aware	32.42	17.62	26.45	21.15	55.79	48.92	52.13	38.45	33.75	25.66	29.16	**65.18**	46.70	54.41
SCE-Net	39.61	37.13	26.88	31.18	61.98	45.99	52.80	39.36	37.42	26.89	31.29	62.03	46.26	52.99
MCN	34.58	33.37	22.64	26.98	61.31	41.67	49.62	34.51	35.02	22.55	27.43	61.34	41.74	49.67
TDRG	35.29	21.39	26.17	23.54	60.71	47.75	53.45	31.78	16.90	20.08	18.35	59.13	38.91	46.93
LAPE	30.53	14.94	20.09	17.14	58.70	38.74	46.67	30.55	15.03	20.10	17.20	59.02	38.95	46.93
Ours	**41.81**	**39.09**	**32.05**	**35.22**	**63.85**	50.54	**56.42**	**41.60**	**39.11**	**32.10**	**35.26**	63.71	**50.86**	**56.57**

Table 3: Comparison of different methods based on Res18 on physical compatibility prediction task.

Method	ACC	NDCG	mAP	CF1	OF1
Ours(wo/msp)	0.6343	0.7110	41.13	35.21	58.85
Ours(wo/smk)	0.8075	0.8411	42.38	38.81	55.69
Ours(wo/msp_smk)	0.6286	0.7159	41.51	31.50	54.15
Ours	0.9309	0.9434	41.60	35.26	56.57

(a)

Num	ACC	NDCG	mAP	CF1	OF1
16	0.8081	0.8887	41.48	39.97	57.03
32	0.5598	0.6731	40.72	33.38	57.41
64	0.9309	0.9434	41.60	35.26	56.57
128	0.8818	0.8818	41.49	31.17	57.96

(b)

Table 4: Results of P-Net under (a) different cases and (b) various user anchor numbers.

3.2.2 How is P-Net for compatibility evaluation regarding physical attributes?

The experimental result is shown in Table 3 where P-Net obtains the best performance in most cases, and we can conclude the following interesting points:

(1) Although the Resnet backbone can obtain competitive performance on metric learning with high ACC and NDCG (as shown in Table 2), the performance of multi-label classification is not as high as we expected. CSN, type-aware and SCE-Net are expected to obtain high metric learning performance as they were initially designed for conditional similarity learning with/without explicit type supervision. However, the results are not as we expected due to the complicated balance of the two tasks. The proposed P-Net obtains better performance in dealing with both of the tasks as it, on the one hand, takes advantage of message passing and style mask strategy for learning advanced feature embeddings for user personality-style mapping, and on the other hand, utilizes a Transformer encoder to capture the potential correlation between the feature embeddings and the physical label embeddings for effective physical label compatibility learning.

(2) LAPE and the proposed P-Net use user anchors for user outfit mapping, and the similarity scores are computed based on the matrix-vector multiplication. The reason why P-Net obtains better performance than LAPE may lie in two aspects: 1) P-Net first captures the relations among different nodes in the outfit graph and then learns the similarity and difference with the style mask, thus the obtained feature embeddings are more discriminant for the personality style comparison; 2) P-Net uses Transformer encoder to encode the feature embeddings that obtained from the feature extractor while LAPE does not take any strategy to enhance the effectiveness of the feature embeddings for the multi-label classification problem. This is reasonable as LAPE is not originally designed to deal with the multi-label classification problem.

3.3 What can affect the performance of P-Net?

(1) From Figure 3, we can know that when $\lambda = 0.01$, P-Net can obtain the best performance. λ is the parameter to balance the importance of the tasks of metric learning and multi-label classification, and thus it should be chosen carefully for the proposed P-Net.

(2) The message-passing network and style mask strategy are important modules of P-Net for user personality-style mapping and it is necessary to explore how they affect the overall performance. Based on this regard, we have three variations: **Ours(wo/msp):** Ours without message passing network. **Ours(wo/smk):** Ours without style mask strategy. **Ours(wo/msp_smk):** Ours without both. The results of the variations are shown in Table 4 (a), from which we can see that the performance of P-Net without message passing or style mask is hard to be satisfactory while one of them can improve the performance, especially on ACC. Compared with the variations, P-Net can obtain the highest ACC and NDCG though the multi-label classification performance is not the highest. This is reasonable as the mes-

sage passing and style mask are incorporated in the personality-style mapping space, so the metric learning performance can be improved. But it does not mean that the two modules can benefit the classification task.

(3) From Table 4 (b), we can see that different numbers of user anchors lead to different performances. The anchor number is not just as small as possible or as large as possible and we use 64 anchors in the experiment.

4 Conclusion

This paper proposes a deep neural model to learn fashion personality corresponding to different MBTI personality traits for user understanding and personalized fashion recommendation. The model also considers users' physical attributes to learn outfit compatibility, with which the recommendation results can fit the users' inner personality traits and physical appearance. The model's effectiveness has been evaluated with qualitative and quantitative results on a new stylish outfit of personality (SOP) dataset. Due to the imbalanced distributions of the outfit's physical labels and the user personalities, the multi-label classification performance of P-Net and the comparison methods is hardly satisfactory for practical use. Additionally, exploring the fashion personality in this paper is still at an initial step. The proposed framework mainly focuses on learning visual image information and it needs an external model to learn the textual information for personality learning, which increases its complexity for implementation.

In the future, we need to, on the one hand, enrich the dataset by collecting more outfit data with balanced physical label distribution and by enhancing user information with larger amount and various personality types, on the other hand, develop a multi-modality learning model to fully consider visual and textual information for effective fashion personality learning and personalized outfit recommendation.

Acknowledgement This research is funded by the Laboratory for Artificial Intelligence in Design (Project Code: RP3-2), Innovation and Technology Fund, Hong Kong Special Administrative Region.

References

[1] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

[2] Isabel Briggs Myers and Peter B Myers. Gifts differing: Understanding personality type. *Davis-Black Publishing*, 1995.

[3] Jinpeng Chen, Yuan Cao, Fan Zhang, Pengfei Sun, and Kaimin Wei. Sequential intention-aware recommender based on user interaction graph. In *Proceedings of the 2022 International Conference on Multimedia Retrieval*, pages 118–126, 2022.

[4] Xu Chen, Hanxiong Chen, Hongteng Xu, Yongfeng Zhang, Yixin Cao, Zheng Qin, and Hongyuan Zha. Personalized fashion recommendation with visual explanations based on multimodal attention network: Towards visually explainable recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 765–774, 2019.

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[6] Jianfeng Dong, Zhe Ma, Xiaofeng Mao, Xun Yang, Yuan He, Richang Hong, and Shouling Ji. Fine-grained fashion similarity prediction by attribute-specific embedding learning. *IEEE Transactions on Image Processing*, 30:8410–8425, 2021.

[7] Weili Guan, Fangkai Jiao, Xuemeng Song, Haokun Wen, Chung-Hsing Yeh, and Xiaojun Chang. Personalized fashion compatibility modeling via metapath-guided heterogeneous graph learning. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 482–491, 2022.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[9] Shintami Chusnul Hidayati and Yeni Anistyasari. Body shape calculator: Understanding the type of body shapes from anthropometric measurements. In *Proceedings of the 2021 International Conference on Multimedia Retrieval*, pages 461–465, 2021.

[10] Chin-Yu Hsieh and Yung-Ming Li. Fashion recommendation with social intelligence on personality and trends. In *2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI)*, pages 85–90. IEEE, 2019.

[11] Waikeung Wong Kaicheng Pang, Xingxing Zou. Dress well via fashion cognitive learning. *arXiv preprint arXiv:2208.00639*, 2022.

[12] Brian Kulis, Mátyás A Sustik, and Inderjit S Dhillon. Low-rank kernel learning with bregman matrix divergences. *Journal of Machine Learning Research*, 10(2), 2009.

[13] Jack Lanchantin, Tianlu Wang, Vicente Ordonez, and Yanjun Qi. General multi-label image classification with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16478–16488, 2021.

[14] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1096–1104, 2016.

[15] Yash Mehta, Samin Fatehi, Amirmohammad Kazameini, Clemens Stachl, Erik Cambria, and Sauleh Eetemadi. Bottom-up and top-down: Predicting personality with psycholinguistic and language model features. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 1184–1189. IEEE, 2020.

[16] Isabel Briggs Myers and Linda K Kirby. Introduction to type: a guide to understanding your results on the myers-briggs type indicator, palo alto, 1998.

[17] Jenny Denise Seidenschwarz, Ismail Elezi, and Laura Leal-Taixé. Learning intra-batch connections for deep metric learning. In *International Conference on Machine Learning*, pages 9410–9421. PMLR, 2021.

[18] Anisha Talwar, Jaspreet Kaur, and Varun Duggal. The impact of myers-briggs personality types (introversion and extroversion) on purchase intention towards international fashion brands of millennials in india. *Academy of Marketing Studies Journal*, 24(2):1–15, 2020.

[19] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.