Neural Compatibility Modeling With Probabilistic Knowledge Distillation

Xianjing Han, Xuemeng Song, Yiyang Yao, Xin-Shun Xu, and Liqiang Nie

***Abstract*— In modern society, clothing matching plays a pivotal role in people's daily life, as suitable outfits can beautify their appearance directly. Nevertheless, how to make a suitable outfit has become a daily headache for many people, especially those who do not have much sense of aesthetics. In the light of this, many research efforts have been dedicated to the task of complementary clothing matching and have achieved great success relying on the advanced data-driven neural networks. However, most existing methods overlook the rich valuable knowledge accumulated by our human beings in the fashion domain, especially the rules regarding clothing matching, like "coats go with dresses" and "silk tops cannot go with chiffon bottoms". Towards this end, in this work, we propose a knowledge-guided neural compatibility modeling scheme, which is able to incorporate the rich fashion domain knowledge to enhance the performance of the compatibility modeling in the context of clothing matching. To better integrate the huge and implicit fashion domain knowledge into the data-driven neural networks, we present a probabilistic knowledge distillation (PKD) method, which is able to encode vast knowledge rules in a probabilistic manner. Extensive experiments on two real-world datasets have verified the guidance of rules from different sources and demonstrated the effectiveness and portability of our model. As a byproduct, we released the codes and involved parameters to benefit the research community.**

***Index Terms*— Multi-modal, compatibility modeling, probabilistic knowledge distillation.**

I. Introduction

ACCORDING to the FashionUnited, the global fashion and apparel industry is valued of three trillion dollars, making up two percent of the world's gross domestic product.[1] The blossom of the fashion market demonstrates people's great demand of clothing. In fact, clothing plays a vital role in people's daily life, since a decent outfit can improve one's appearance right away. Nevertheless, the tremendous fashion items (e.g., the tops and bottoms) tend to get people overwhelmed and make it rather intractable for them to compose suitable outfits, especially for those who lack the good taste of clothing matching. Fortunately, with the proliferation of many online fashion communities (e.g., Ssense[2] and Chictopia,[3]) plenty of well-composed outfits shared by fashion experts, as shown in Figure 1, are made publicly available. In fact, the rich real-world outfit data has facilitated many researchers to investigate the solution for automatic clothing matching.

Fig. 1. Examples of outfit compositions.

The great success of deep learning methods in representation learning has propelled researchers to be keen on tackling the clothing matching problem with advanced deep neural networks to learn powerful representations for fashion items. As pure data-driven methods, neural networks heavily rely on large amounts of labeled data; whereas they often overlook the value of human knowledge and suffer from the poor interpretability. As a matter of fact, the cognitive process of human beings enables us to learn from not only the concrete examples but also the general knowledge, which can be derived from the rich human experiences. Since clothing matching has become an essential aspect of people's daily life, it has accumulated a variety of valuable knowledge, like the matching rule "tank tops go better with shorts instead of dresses". Although certain knowledge rules may be subjective to some extent, most of them have been widely accepted by the public as common sense. Therefore, to get rid of the defects of the deep neural networks and improve the matching performance, in this work, we work towards leveraging the domain knowledge to boost the performance of data-driven clothing matching.

Manuscript received January 23, 2019; revised June 4, 2019 and July 15, 2019; accepted August 6, 2019. Date of publication August 27, 2019; date of current version October 9, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61772310, Grant 61702300, Grant 61702302, Grant 61802231, and Grant U1836216, in part by the Project of Thousand Youth Talents 2016, in part by the Shandong Provincial Natural Science and Foundation under Grant ZR2019JQ23 and Grant ZR2019QF001, and in part by the Future Talents Research Funds of Shandong University under Grant 2018WLJH63. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dong Tian. *(Corresponding authors: Xuemeng Song; Liqiang Nie.)*

X. Han, X. Song, and L. Nie are with the School of Computer Science and Technology, Shandong University, Qingdao 266237, China (e-mail: hanxianjing2018@gmail.com; sxmustc@gmail.com; nieliqiang@gmail.com).

Y. Yao is with the State Grid Zhejiang Electric Power Co., Ltd., Hangzhou 310007, China (e-mail: yao_yiyang@zj.sgcc.com.cn).

X.-S. Xu is with the School of Software, Shandong University, Jinan 250101, China (e-mail: xuxinshun@sdu.edu.cn).

Digital Object Identifier 10.1109/TIP.2019.2936742

[1] https://fashionunited.com/global-fashion-industry-statistics

[2] https://www.ssense.com/

[3] http://www.chictopia.com/

Without loss of generality, the problem we pose here can be cast as the compatibility modeling between complementary fashion items, such as the top and bottom. However, modeling the compatibility between fashion items from both data-driven and knowledge-driven perspectives is challenging due to the following reasons. 1) The human knowledge on fashion domain is usually unstructured and fuzzy as it tends to be implicitly conveyed by the large amount of outfit compositions of fashion experts, making it difficult to be directly employed by neural networks. Therefore, it is an arduous task to establish a set of structured knowledge rules based on the vast fuzzy domain knowledge for the clothing matching. 2) How to properly encode such knowledge rules into the pure data-driven learning scheme and enable the model to learn from both the specific data and the general rules poses another challenge for us. And 3) verifying the portability of utilizing the domain knowledge in the context of clothing matching across different datasets remains largely untapped.

To address the above challenges, we propose a neural compatibility modeling scheme with probabilistic knowledge distillation based on the Bayesian Personalized Ranking (BPR) [1], dubbed PKD-DBPR, as shown in Figure 2, which is able to learn from both the specific data samples and the general fashion domain knowledge. In particular, we present a teacher-student scheme, where the teacher network is to guide the training process of the student network with the domain knowledge regularization. Namely, the student network is encouraged to not only achieve good performance in the compatibility modeling but also emulate the knowledge-regularized teacher network well. As a pure data-driven learning model, the student network is devised as a dual-path neural network for the purpose of learning a latent compatibility space to unify the complementary fashion items from heterogeneous spaces. In order to comprehensively model the compatibility, the student network seamlessly integrates the visual and contextual modalities of fashion items by imposing hidden layers over the concatenations of their representations on different modalities. Ultimately, towards the compatibility modeling, we adopt the pairwise preference between complementary fashion items and hence build our student network based upon the BPR framework.

As to the teacher network construction for the knowledge distillation, we introduce a novel knowledge encoding method, probabilistic knowledge distillation (PKD), an extension of our previous work [2]. In [2], we proposed an attentive knowledge distillation method (AKD), which is able to encode the manually-screened fashion knowledge rules into the teacher network regularizers to guide the student network, where the attention mechanism [3] is adopted to adaptively assign the knowledge rule confidences. Motivated by the fact that the human knowledge on fashion domain can be vast and fuzzy, making it infeasible to manually derive the matching rules and learn the corresponding rule confidences, we devise the PKD to fulfil the knowledge encoding task from the probabilistic perspective. In particular, PKD is capable of coping with the abundant domain knowledge without the manual screening efforts and thus able to incorporate more thorough domain knowledge than AKD. Moreover, in a sense, the nature of PKD that expresses the knowledge rules in a probabilistic manner alleviates the trouble of rule confidence learning. Notably, both AKD and the proposed PKD aim to harness the neural networks with the rich fashion domain knowledge and enhance the interpretability for the compatibility assessment of a given item pair with the help of the rules extracted from the domain knowledge. Beyond that, to gain a comprehensive understanding of the rule guidance, the knowledge rules utilized in this work are derived from not only our training (internal) dataset but also the publicly available (external) knowledge. The main contributions of this work can be summarized in threefold:

- We present a compatibility modeling scheme with probabilistic knowledge distillation in the context of clothing matching, which enables the scheme to learn from not only the specific data samples but also the general domain knowledge. The proposed PKD facilitates the scheme to incorporate more abundant rules and dispense with the trouble of rule confidence learning, as compared to our previous AKD.
- To get a thorough understanding of the rule guidance in PKD-DBPR, we explore both the internal and external knowledge rules that can be extracted from our own training dataset and the external public knowledge, respectively.
- Extensive experiments conducted on two real-world datasets FashionVC and ExpFashion demonstrate the portability and effectiveness of the proposed scheme. As a byproduct, we released the codes and parameters to benefit other researchers.[4]

The rest of the paper is organized as follows. Section II briefly reviews the related work. In Section III, we introduce the proposed probabilistic knowledge distillation. The experimental results and analyses are presented in Section IV, followed by the conclusion and future work in Section V.

II. Related Work

A. Fashion Analyses

In recent years, the huge potential of the fashion market has attracted increasing attention of researchers from various research communities. Existing researches mainly focus on clothing retrieval [4], [5], fashion trending prediction [6], fashionability prediction [7] and compatibility modeling [2], [8], [9]. For example, Liu *et al.* [5] presented a latent Support Vector Machine [10] model for both occasion-oriented outfit and item recommendation based on a dataset of manually annotated wild street photos. Because of the infeasibility of human annotated dataset, some researchers have resorted to other sources, where real-world data can be acquired automatically. For example, McAuley *et al.* [11] introduced a general framework to model the human visual preference for a given pair of objects based on a real-world dataset of co-purchase products on Amazon. In particular, they extracted visual features with convolutional neural networks (CNNs) and utilized a similarity metric to model the human notion of complementary objects.

[4] https://tinyurl.com/y7pftrj7/

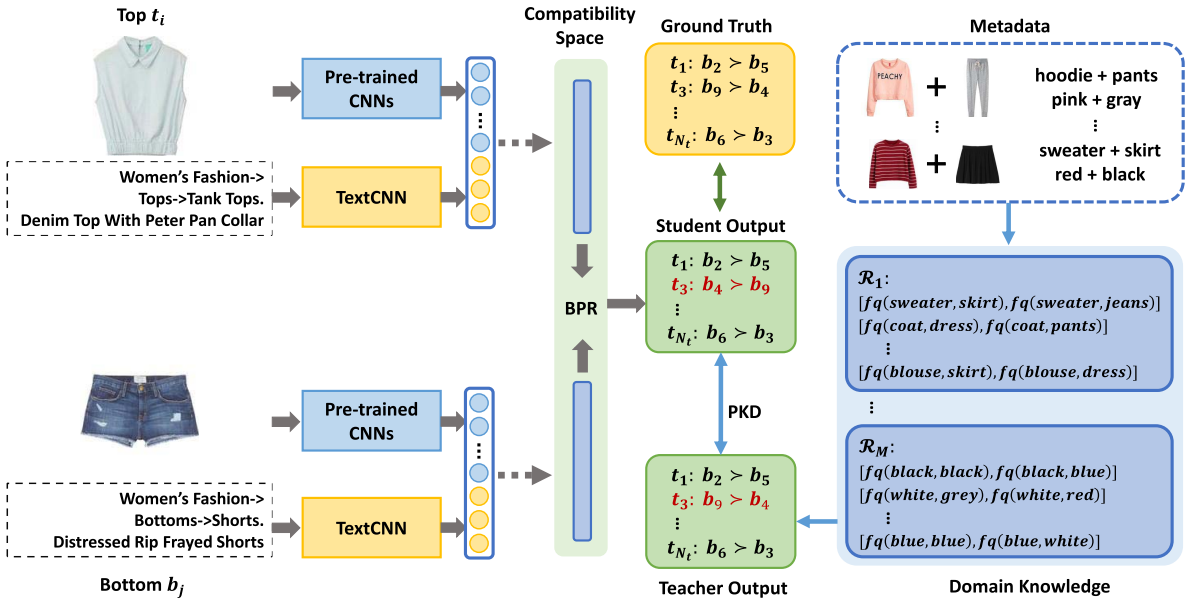


Fig. 2. Illustration of the proposed scheme. The student network, comprising dual-path neural networks, aims to learn the latent compatibility space where the implicit preference among items can be modeled via Bayesian Personalized Ranking (BPR). The teacher network compiles the domain knowledge by PKD and guide the student network to achieve the knowledge distillation. t_i: top; b_j: bottom; "$\succ$": pairwise preference; "->": the category hierarchy; $\mathcal{R}_m$: the m-th matching rule set adopted in PKD; $fq(\cdot)$: the frequency of the attribute value pair.

Likewise, He and Mcauley [12] proposed a scalable matrix factorization approach that incorporates the visual features of product images to fulfil the recommendation task. However, former researches on fashion analysis mainly investigate the practical problem on the visual data but ignored the value of contextual information of fashion items. Towards this end, Song *et al.* [2] comprehensively investigated the problem of complementary fashion item matching with a multi-modal fashion dataset, FashionVC, collected from Polyvore.[5] Later, Lin *et al.* [9] explored user comments to improve the fashion recommendation quality, where a dataset, ExpFashion, comprising not only the multi-modal data of fashion items but also the user comments, is created. Although existing efforts have obtained remarkable achievements, these studies mainly focus on modeling the compatibility purely based on the data-driven deep learning methods but overlook the value of human knowledge. Distinguished from existing researches, we aim to employ the fashion domain knowledge to guide the pure data-driven neural networks, and hence reduce the model reliance on the large amounts of labeled data and improve the model interpretability as a side product.

B. Knowledge Distillation

Deep neural networks have achieved distinguished performance in various application domains ranging from natural language processing [13], [14] to computer vision [15], [16]. To boost the learning performance, one common way in machine learning domain is to ensemble multiple models and average the predictions. However, it is intractable to ensemble the large neural networks due to its tremendous computational expense and cumbersome deployment. Towards this end, in 2015, Hinton *et al.* [17] first introduced a knowledge distillation framework to transfer the knowledge from a large cumbersome model to a small model, which facilitates the model deployment. Later, the knowledge distillation framework is adopted to accelerate the training process of the neural network [18] and improve the model portability [19], [20].

Similarly, inspired by the knowledge distillation framework, Hu *et al.* [21] introduced an iterative teacher-student distillation approach. The approach can be intuitively explained in analogous to the human education where the teacher is aware of systematic general rules and the student learns from the teacher by iteratively imitating the teacher's solutions to specific questions. In particular, they equipped the teacher neural network with regularizations that encode the domain knowledge, represented by logic rules, to tackle various natural language processing tasks. The domain knowledge not only boosts the model performance, but also improves the interpretability of the pure data-driven model. In addition, Yu *et al.* [22] studied the guidance of both the internal and external linguistic knowledge in the context of visual relationship detection, and the experimental results are promising. Furthermore, Alashkar *et al.* [23] introduced a makeup recommendation and synthesis system, where both the makeup art domain knowledge and makeup expert experience are incorporated into the neural network to boost the performance of makeup recommendation. Although the knowledge distillation in deep neural networks has been successfully applied to solve the visual relationship detection [22], sentence sentiment analysis [21] and name entity recognition [24], limited efforts have been dedicated to the fashion domain. Towards this end, in this work, we aim to devise a proper knowledge encoding method to take advantage of the fashion domain knowledge as a guidance in the traditional neural models.

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

III. Neural Compatibility Modeling

A. Notation

To improve the readability, we first declare some notations used in this paper. We use bold capital letters (e.g., $\mathbf{X}$) and bold lowercase letters (e.g., $\mathbf{x}$) to represent matrices and vectors, respectively. Let the non-bold letters (e.g., x) denote scalars and Greek letters (e.g., β) denote parameters. The vectors without clarification are in the column forms. In addition, we use $||\mathbf{A}||_F$ to represent the Frobenius norm of matrix $\mathbf{A}$.

B. Problem Formulation

To make a harmonious outfit, people prefer to choose clothes with high compatibility, such as "a short denim jacket plus a corded lace and crepe dress" or "a striped cashmere sweater plus skinny jeans". In this work, we focus on tackling the essential problem of compatibility modeling for clothing matching.

Formally, we have a set of tops $\mathcal{T} = \left\{t_1, t_2, \cdots, t_{N_t}\right\}$ and bottoms $\mathcal{B} = \left\{b_1, b_2, \cdots, b_{N_b}\right\}$, where N_t and N_b represent the number of the tops and bottoms, respectively. We use $\mathbf{v}_i^t\,(\mathbf{v}_i^b) \in \mathbb{R}^{D_v}$ and $\mathbf{c}_i^t\,(\mathbf{c}_i^b) \in \mathbb{R}^{D_c}$ to respectively denote the visual and contextual embeddings of t_i (b_j), where D_v and D_c stand for the dimensions of the corresponding embeddings. In addition, each fashion item is characterized by a set of attributes (e.g., the *color* and *category*) $\mathcal{A} = \{a_m\}_{m=1}^M$, where a_m is the m-th attribute and M is the total number of attributes. For each attribute a_m, we also have a set of distinct values $\mathcal{E}_m = \left\{val_m^1, val_m^2, \cdots, val_m^{E_m}\right\}$, where E_m is the total number of the values. For example, the values for the attribute *color* include *blue*, *white* and *red*. Meanwhile, we have a set of positive top-bottom pairs $\mathcal{S} = \left\{(t_{i_1}, b_{j_1}), (t_{i_2}, b_{j_2}), \cdots, (t_{i_N}, b_{j_N})\right\}$ obtained from the dataset that consists of the compositions of fashion experts, where N denotes the total number of positive pairs. Accordingly, for each top t_i, a set of positive bottoms $\mathcal{B}_i^+ = \left\{b_j \in \mathcal{B} | (t_i, b_j) \in \mathcal{S}\right\}$ can be derived.

Let m_{ij} stand for the compatibility between top t_i and bottom b_j, based on which we thus can derive a ranking list of bottoms b_j's for a given top t_i and address the practical clothing matching problem. In this work, to accurately measure m_{ij}, we devote to devise a neural compatibility modeling scheme, which is capable of utilizing the general knowledge rules to supervise the training of the data-driven model and hence boost the performance. Table I summarizes the main notations used in this work.

C. Data-Driven Compatibility Modeling

In fact, due to the heterogeneity of complementary fashion items, it is not advisable to directly measure their compatibility from the original feature spaces. To bridge this gap, we argue that there should be a latent compatibility space where the compatibility between complementary items can be well measured. Moreover, the fact that the compatibility can be always affected by sophisticated factors, ranging from the color and style to material and pattern, propels us to learn the latent space in the non-linear manner. Naturally, we employ the neural network to explore the latent compatibility space for its superior performance in various machine learning tasks.

TABLE I
Summary of the Main Notations

Notation	Explanation
t_i	The i-th top.
b_j	The j-th bottom.
$\mathcal{S}$	The set of positive top-bottom pairs.
a_m	The m-th attribute.
$val_m^{t_i}(val_m^{b_j})$	The value regarding attribute a_m of t_i (b_j).
$\tilde{\mathbf{z}}_i^t(\tilde{\mathbf{z}}_i^b)$	The latent representation of t_i (b_j).
m_{ij}	The compatibility between t_i and b_j.
$\mathcal{F}_m$	The set of attribute value pairs with their co-occurrence frequency.
$\mathcal{R}_m$	The rule set of the m-th attribute.
Θ	The to-be-learned set of parameters.

In fact, each fashion item usually possesses multiple modalities, such as the visual and contextual modalities, which complementarily characterize the same fashion item. For example, the color and shape of the fashion items can be intuitively reflected by the visual modality, while the category and material information can be concisely represented by the contextual modality. To fully exploit the rich multi-modal data of fashion items, we resort to the multi-layer perceptron (MLP), which can model the semantic relation between different modalities of fashion items. In particular, we deploy K hidden layers over the concatenation of the visual and contextual representations as follows,

$$\begin{cases} \mathbf{z}_{i0}^x = \begin{bmatrix} \mathbf{v}_i^x \\ \mathbf{c}_i^x \end{bmatrix}, \\ \mathbf{z}_{ik}^x = s(\mathbf{W}_k^x \mathbf{z}_{i(k-1)}^x + \mathbf{b}_k^x), \quad k = 1, \cdots, K, \ x \in \{t, b\}, \end{cases} \tag{1}$$

where $\mathbf{z}_{ik}^x$ is the hidden representation, $\mathbf{W}_k^X$ and $\mathbf{b}_k^x$ are weight matrices and biases, respectively. t and b denote *top* and *bottom*. $s : \mathbb{R} \mapsto \mathbb{R}$ is a non-linear function applied element wise and we choose the sigmoid function $s(x) = \frac{1}{1+e^{-x}}$ in this work. The latent representation of the fashion item is defined as the output of the K-th layer, i.e., $\tilde{\mathbf{z}}_i^x = \mathbf{z}_{iK}^x \in \mathbb{R}^{D_l}, x = \{t, b\}$, 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,

$$m_{ij} = (\tilde{\mathbf{z}}_i^t)^T \tilde{\mathbf{z}}_j^b. \tag{2}$$

In a sense, we can safely argue that the top-bottom pairs composed together by fashion experts are the positive (compatible) samples. However, it may be too absolute to claim that the non-composed fashion item pairs are the negative (incompatible) ones, due to that they can be the missing potential positive pairs whose items may be composed together later. In order to model the implicit relations between the tops and bottoms, we adopt the BPR framework [1], which has shown excellent performance in the implicit preference modeling [25]–[27]. In addition, we argue that as for top t_i,

bottoms in the positive set $\mathcal{B}_i^+$ are more compatible than the non-composed 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 compared with bottom b_k, bottom b_j is more compatible with top t_i. Then according to [1], the objective function can be written as follows,

$$\mathcal{L}_{bpr} = \sum_{(i,j,k)\in\mathcal{D}_S} -ln(\sigma(m_{ij} - m_{ik})) + \frac{\lambda}{2}\|\Theta\|_F^2, \quad (3)$$

where λ is the non-negative hyperparameter to avoid overfitting. Θ denotes the set of parameters (i.e., $\mathbf{W}_k^x$ and $\mathbf{b}_k^x$).

D. Probabilistic Knowledge Distillation

As an indispensable part of people's daily life, clothing matching domain has accumulated considerable human knowledge. For example, it is widely recognized that sweaters go better with jeans than shorts, while a silk top can hardly go with a knit bottom. To take full advantage of the valuable domain knowledge, we employ the knowledge distillation technique to guide the neural networks and thus enable the model to learn from not only the specific data but also the general knowledge rules. In particular, we adopt the teacher-student scheme [21], which shares the same underlying philosophy with the human education, where the teacher equipped with the professional knowledge regularization can guide students with his/her solutions to specific problems. In particular, apart from achieving the outstanding prediction performance, the data-driven student network p is also encouraged to imitate the behaviour of the teacher network q. Accordingly, the objective function at iteration t can be written as follows,

$$\Theta^{t+1} = \arg\min_{\Theta} \sum_{(i,j,k)\in\mathcal{D}_S} \Big\{(1-\rho)\mathcal{L}_{bpr}(m_{ij}^p, m_{ik}^p) + \rho\mathcal{L}_{crs}(q^{(t)}(i, j, k), p(i, j, k)\Big\} + \frac{\lambda}{2}\|\Theta\|_F^2, \quad (4)$$

where $\mathcal{L}_{crs}$ represents the cross-entropy loss. $p(i, j, k)$ and $q(i, j, k)$ are the sum-normalized distributions over the compatibility scores predicted by the student network p and teacher network q (i.e., $[m_{ij}^p, m_{ik}^p]$ and $[m_{ij}^q, m_{ik}^q]$), respectively. ρ is the imitation parameter calibrating the relative weights of these two terms.

Considering that the human knowledge regarding fashion can be vast and fuzzy, making it intractable to manually screen the large amount of matching rules and set the rule confidences, we encode the matching rules into the teacher network in a probabilistic manner. Regarding the matching rule derivation, for each attribute a_m, we can obtain a set of value pairs with their co-occurrence frequency $\mathcal{F}_m$ in the dataset, which is defined as follows,

$$\Big\{(val_m^{(t)}, val_m^{(b)}) : fq(val_m^{(t)}, val_m^{(b)})|val_m^{(t)}, val_m^{(b)} \in \mathcal{E}_m\Big\}, \quad (5)$$

where $val_m^{(t)}$ and $val_m^{(b)}$ represent the values of top and bottom regarding the m-th attribute. $fq(val_m^{(t)}, val_m^{(b)})$ refers to the co-occurrence frequency of the value pair $(val_m^{(t)}, val_m^{(b)})$, and we calculate $fq(val_m^{(t)}, val_m^{(b)}) =$

$$\sum_{t_i\in\mathcal{T}, b_j\in\mathcal{B}_i^+} I(val_m^{(t)} = val_m^{(t_i)} \wedge val_m^{(b)} = val_m^{(b_j)}), \quad (6)$$

where $val_m^{(t_i)}$ and $val_m^{(b_j)}$ denote the values regarding attribute a_m of top t_i and bottom b_j in $\mathcal{S}$, respectively. $I(\cdot)$ is the indicator function, which takes on a value of 1 if and only if its argument is true, and 0 otherwise. In a sense, $fq(val_m^{(t)}, val_m^{(b)})$ can be treated as the confidence of the rule regarding $(val_m^{(t)}, val_m^{(b)})$. The details pertaining to the construction of the rule set $\mathcal{R}_m$ from $\mathcal{F}_m$ will be introduced later in the following Subsection III-E.

As for the teacher network construction, on the one hand, we expect that the student network p can learn well from the teacher network q, and hence adopt the closeness between the compatibility prediction of these two networks. On the other hand, we propose to utilize the rule regularizers to encode the general domain knowledge. Accordingly, we adapt the teacher network construction method proposed in [21], [28] as follows,

$$\min_{\mathbf{q}} KL(q(i, j, k)||p(i, j, k)) - C\sum_m \mathbb{E}_q[\mathbf{g}_m(i, j, k)]. \quad (7)$$

Accordingly, we have the following closed-form solution,

$$q^*(i, j, k) \propto p(i, j, k)\exp\left\{\sum_m C g_m(i, j, k)\right\}, \quad (8)$$

where $g_m(i, j, k)$ is the m-th attribute rule constraint function introduced to reward or penalize the student network in a probabilistic manner. In particular, we define $g_m(i, j, k) =$

$$\begin{cases} h([fq(val_m^{(i)}, val_m^{(j)}), fq(val_m^{(i)}, val_m^{(k)})]), & if \begin{cases}\tau_m(ij)=1,\\ \tau_m(ik)=1,\end{cases}\\ [0, 0], & others, \end{cases} \quad (9)$$

where $val_m^{(a)}$ denotes the value regarding attribute a_m of sample a and $\tau_m(ab) = 1$ means that the value pair $(val_m^{(a)}, val_m^{(b)})$ belongs to $\mathcal{R}_m$. h is the sum-normalization function $h([u, v]) = [\frac{u}{u+v}, \frac{v}{u+v}]$, which is able to cast the co-occurrence frequency to the probabilistic representation. The workflow of PKD is illustrated in Figure 3.

Notably, the teacher network is constructed from the student network in the initial stage, which may result in the poor guidance at the beginning of the training process. Therefore, we expect the whole framework favors to the prediction of the ground truth more at first and gradually biases towards emulating the teacher network to distill the knowledge. We thus adopt the strategy in [21] that assigns ρ dynamically to keep ρ increasing as the training process goes.

E. Rule Construction

In this work, we aim to utilize the explicit structured matching rules to guide the neural network and hence boost the performance of clothing matching. In general, the compatibility between fashion items is mainly affected by five

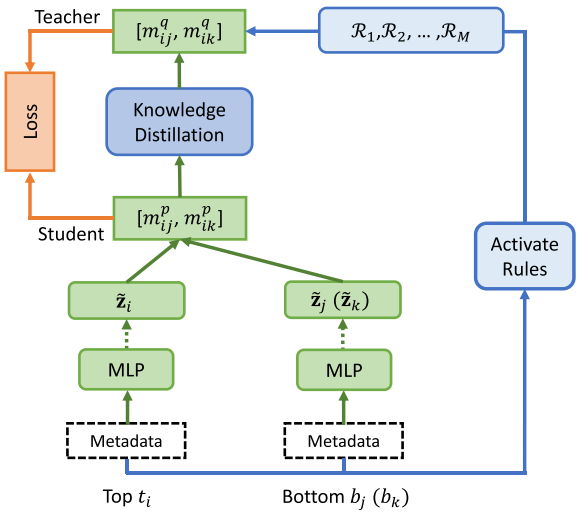


Fig. 3. Illustration of the proposed PKD method.

TABLE II
VALUE EXAMPLES OF EACH ATTRIBUTE

Attribute	Value Example
Color	white, black, green, red, gray, blue
Material	knit, silk, leather, cotton, fur, cashmere
Pattern	pure, grid, dot, floral, number (letter)
Category	coat, dress, skirt, sweater, jeans, hoodie
Brand	Yoin, H&M, Gucci, River Island

attributes: *color*, *material*, *pattern*, *category* and *brand*. Therefore, we take the annotation details in [29] as a reference and define a dictionary with all the possible values of each attribute based on our training dataset. Table II shows several value examples of each attribute. According to Eqn. (5), we conduct the statistics on the co-occurrence of attribute value pairs and obtain the $\mathcal{F}_m$, based on which we can easily acquire the matching rules for PKD. In particular, we employ the high co-occurrence value pairs and low ones in $\mathcal{F}_m$ to constitute the final co-occurrence value pair set $\mathcal{R}_m$ for PKD. The underlying philosophy is that the attribute value pairs with high co-occurrence frequency can be treated as the positive rules that people should follow in outfit composition, while those with low co-occurrence frequency can be regarded as the negative rules that people should avoid.

To make the rules more intuitive, we use $[fq(value1, value2), fq(value1, value3)]$ represents the rule in PKD. For example, $[fq(coat, dress), fq(coat, skirt)]$ stands for the rule regarding the matching between coats and dress/skirts. Moreover, according to Eqn. 9, our model needs to judge whether the given fashion item pair activates certain rule. Here we define that the given pair triggers the rule $[fq(value1, value2), fq(value1, value3)]$, if and only if the *value1*, *value2* and *value3* agree with the attribute values, extracted from the metadata, of the given top and two bottom candidates, respectively.

IV. EXPERIMENT

To verify the portability and effectiveness of the proposed model, we conducted extensive experiments on two real-world datasets FashionVC and ExpFashion. In this section, we first introduce the experimental settings in Subsection IV-A and then present the results of each experiment in the following subsections. In particular, in Subsection IV-B, we compare the proposed PKD-DBPR with several traditional compatibility modeling methods to verify the effectiveness of our PKD-DBPR. To gain more deep insights on knowledge encoding methods, we further make the detailed comparison between AKD and PKD in Subsection IV-C. Besides, we analyze the internal and external rule guidance of PKD-DBPR in Subsection IV-D, and justify its practical value in the application of complementary fashion item retrieval in Subsection IV-E.

A. Experimental Settings

1) Dataset: In this work, to evaluate our model, we adopted two real-world datasets FashionVC [30] and ExpFashion [9], both of which are collected from the online fashion community Polyvore. **FashionVC** consists of 20,726 outfits with 14,871 tops and 13,663 bottoms, while **ExpFashion** is comprised of 200,745 outfits with 29,113 tops and 20,902 bottoms. Each fashion item in FashionVC and ExpFashion is associated with the visual image, relevant categories and the title description. Pertaining to the derivation of the matching rules for PKD, we first extracted the attribute values of each fashion item in our positive top-bottom pairs based on their visual and contextual metadata. In particular, for attributes (e.g., *material*, *pattern*, *category* and *brand*) that are usually conveyed by the contextual information, we extracted the attribute values by keyword detection. As for the attribute *color*, we directly resorted to the histogram calculation in the HSV space to acquire the color value of each item.

2) Contextual Representation: In this work, we took the title and category labels in different granularity as the contextual description of a fashion item. To obtain the effective contextual representation, instead of using traditional linguistic features [31], [32], we adopted the CNN architecture [33], which has demonstrated its effectiveness in many natural language processing tasks [13], [14]. In particular, we first transformed each contextual description into a concatenated word vector, where each row represents one constituent word and each word is allocated with a publicly available 300-D word2vec [34] vector. Then, we deployed the single channel CNN, which consists of a convolutional layer on top of the concatenated word vectors and a max pooling layer. In particular, we utilized four kernels with the sizes of 2, 3, 4, and 5. For each kernel size, we adopted 100 feature maps and the rectified linear unit (ReLU) as the activation function. Finally, we represented the contextual modality of each item with a 400-D vector.

3) Visual Representation: As for the visual modality, we applied the deep CNNs, which have achieved compelling performance in the image representation learning [35]–[37]. In particular, we chose the pre-trained ImageNet deep neural network provided by the Caffe software package [38], comprising 5 convolutional layers and 3 fully-connected layers. We adopted the output of the fc7 layer as the visual representation. Ultimately, we obtained a 4096-D visual representation for each item.

Regarding the experimental setting, we divided the positive top-bottom pair set $\mathcal{S}$ into the training set $\mathcal{S}_{train}$ (80%), validation set $\mathcal{S}_{valid}$ (10%), and testing set $\mathcal{S}_{test}$ (10%). 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 employed the area under the ROC curve (AUC) [39], [40] as the evaluation metric and adopted the stochastic gradient descent (SGD) [41] with the momentum factor as 0.9 for optimization.

In addition, to determine the optimal values for the regularization parameters (i.e., λ, C), we employed the grid search strategy among the values $\{10^r | r \in \{-4, \cdots, -1\}\}$ and [2, 4, 6, 8], respectively. Furthermore, the mini-batch size, the number of hidden units and learning rate were searched in [32, 64, 128, 256], [128, 256, 512, 1024], and [0.005, 0.01, 0.02, 0.05], respectively. We fine-tuned the proposed model for 40 epochs with the performance on the testing set reported. We empirically found the proposed model achieves the optimal performance with $K = 1$ hidden layer of 1024 hidden units.

B. Comparison of Approaches

Due to the sparsity of our dataset, the matrix factorization based methods [42]–[44] are not much suitable. We thus adopted the following content-based baselines to evaluate the proposed PKD-DBPR.

- **POP**: We used the "popularity" of bottom b_j to measure its compatibility with top t_i. In this work, 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 between fashion items.
- **IBR**: We adopted the image-based recommendation method proposed by [11], which aims to model the compatibility between objects simply based on their visual appearance. In particular, a linear latent style space is learned to facilitate the retrieval of correlated objects with the traditional nearest-neighbor search.
- **ExIBR**: We chose the extension of IBR introduced by [2] as one baseline, where both the visual and contextual data of fashion items are utilized to find the latent style space.
- **Bi-LSTM**: We chose the bidirectional LSTM model in [45] 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 two items: a top and a bottom.
- **BPR-DAE**: We selected the content-based neural scheme introduced by [30], which jointly exploits the implicit preference among items via a dual autoencoder network and the coherent relation between the visual and contextual modalities of fashion items.
- **DBPR**: To get a better understanding of our model, we introduced the baseline DBPR, which is the derivation of our model that removes the guidance of the teacher network and only relies on the student network.
- **AKD-DBPR**: To better evaluate the knowledge encoding method, we introduced the baseline AKD-DBPR [2], which also utilizes the teacher-student scheme to encode the knowledge rules. Differently, AKD-DBPR manually screens the matching rules and assigns the rule confidences with the attention mechanism.

TABLE III
COMPARISON AMONG DIFFERENT APPROACHES IN TERMS OF AUC (%) ON FASHIONVC AND EXPFASHION

Approach	FashionVC	ExpFashion
POP	42.06	48.16
RAND	50.94	50.22
IBR	60.75	66.11
ExIBR	70.33	73.94
Bi-LSTM	66.29	67.17
BPR-DAE	76.16	79.12
DBPR	77.04	85.59
AKD-DBPR-p	78.43	87.27
AKD-DBPR-q	78.52	87.15
PKD-DBPR-p	**78.81**	**88.19**
PKD-DBPR-q	78.67	**88.19**

Since we can choose either the distilled student network p or the teacher network q with a final projection for the testing, we introduced two derivations for AKD-DBPR and PKD-DBPR respectively: AKD-DBPR-p, AKD-DBPR-q, PKD-DBPR-p and PKD-DBPR-q. Here the suffixes "-p" and "-q" refer to utilizing the final student network and teacher network to get the compatibility between fashion items, respectively.

Table III shows the performance comparison among different approaches on two datasets. Notably, regarding the ExpFashion, we randomly sampled 20,000 positive outfits instead of using the whole dataset. From this table, we have the following observations.

1) DBPR outperforms all the other state-of-the-art pure data-driven baselines, which demonstrates the superiority of the proposed content-based neural networks for the compatibility modeling.

2) AKD-DBPR and PKD-DBPR, exploiting the matching rules derived from the training dataset, both surpass DBPR, which confirms the benefit of the knowledge distillation in the context of compatibility modeling. To gain a better understanding of the impact of the knowledge rule guidance, we particularly illustrated the comparison between PKD-DBPR and DBPR on several testing triplets in Figure 4. As we can see, PKD-DBPR performs especially better in cases that the given two bottoms b_j and b_k both seem to be visually compatible to the top t_i, and the domain knowledge can be helpful in distinguishing the more compatible one. Meanwhile, we noticed that incorporating the knowledge rules can also result in certain failed triplets. This can be explained by the fact that certain probabilistic matching rules extracted from the training dataset can be less robust and not applicable for all cases. For example, the rule "the coat goes better with the dress than skirt" adopted by PKD is unsuitable for the first failed triplet misjudged by PKD-DBPR, where the coat looks more compatible with the given skirt.

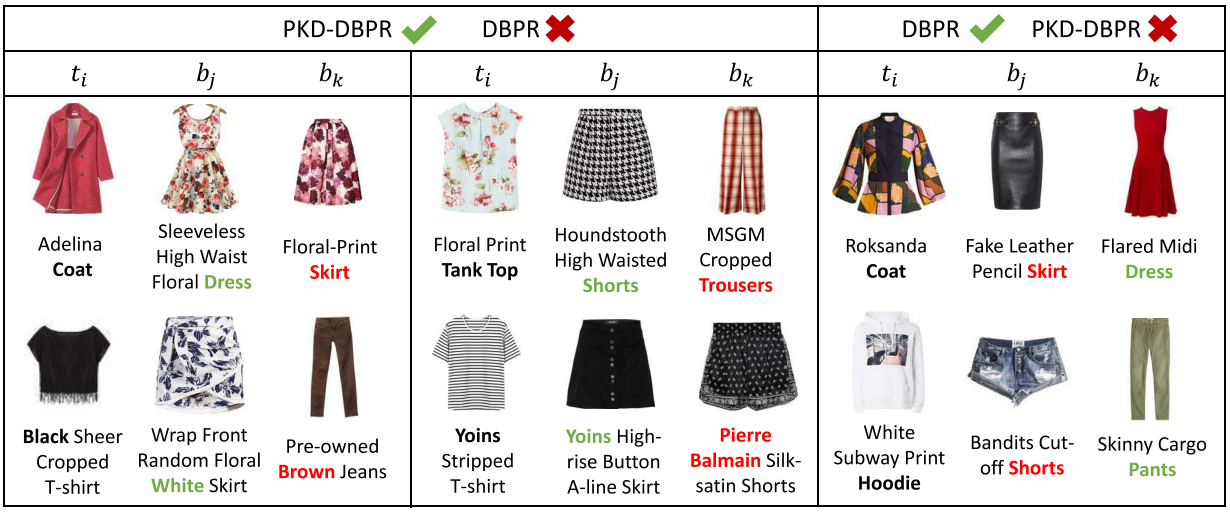


Fig. 4. Comparison between PKD-DBPR and DBPR on the testing triplets. All the triplets satisfy the ground truth that $t_i : b_j \succ b_k$, where "$\succ$" denotes the pairwise preference. We only list the keywords of the metadata of items and bold the values of the rules in PKD-DBPR. For each triplet, the green value refers to the higher co-occurrence frequency with the the black value of top t_i, compared with the red one. For example, the rule activated by the first sample is "coats go better with dresses than skirts".

TABLE IV
COMPARISON BETWEEN AKD-DBPR AND PKD-DBPR ON FASHIONVC AND EXPFASHION IN TERMS OF AUC (%)

Approach	FashionVC		ExpFashion	
	Category	Color	Category	Color
DBPR	77.04	77.04	85.59	85.59
AKD-DBPR-p	78.18	77.57	87.00	86.12
AKD-DBPR-q	78.18	77.77	86.90	86.10
PKD-DBPR-p	**78.36**	77.95	**88.19**	87.71
PKD-DBPR-q	78.24	**77.98**	87.73	**87.75**

3) PKD-DBPR shows superiority over AKD-DBPR, indicating that the advantage of performing knowledge distillation in a probabilistic manner. One plausible explanation is that PKD-DBPR is more capable of handling the vast and fuzzy fashion domain knowledge than AKD-DBPR, and covers more domain knowledge to assist the compatibility modeling.

C. Different Knowledge Distillation Methods

To gain more deep understanding of different knowledge distillation methods, we further carried out detailed comparison between AKD-DBPR and PKD-DBPR. Table IV shows the performance comparison between PKD-DBPR and AKD-DBPR with different rule configurations on two datasets. Each rule configuration constrains the attributes of fashion items that can be used to derive the knowledge rules for AKD-DBPR and PKD-DBPR. In particular, we chose the most essential attributes of fashion items contributing to the clothing matching: *category* and *color*. Notably, according to [2], the rules utilized in AKD-DBPR are manually selected from $\mathcal{F}_m$ by the fashion-lovers. From Table IV, we found that PKD-DBPR consistently outperforms AKD-DBPR with different rule configurations across different datasets. This reconfirms the effectiveness of compiling the knowledge in the probabilistic manner, where the large amount of fuzzy human knowledge can be encoded properly to boost the performance. Furthermore, we observed that rules pertaining to *category* are more robust than those regarding *color* in both AKD-DBPR and PKD-DBPR. The possible reasons are threefold. 1) The category-related knowledge rules, like "the T-shirt goes well with the shorts", are more common and easier to be recognized by the public, presenting the higher robustness and hence providing the better guidance to the data-driven model. 2) The color-related matching rules can be fuzzy and highly subjective, making it hard to properly encode the underlying knowledge. And 3) the category metadata of fashion items is better structured as compared to the color attribute extracted from the visual metadata.

To intuitively reflect the advantage of our PKD, we illustrated the result comparison between AKD-DBPR and PKD-DBPR on several testing triplets in Figure 5. For better illustration, we defined that "*value1* + *value2*" denotes the positive rule in AKD-DBPR, while "no *value1* + *value2*" represents the negative rule. For example, "coat + dress" stands for the positive rule "coats can go with dresses", and "no silk + chiffon" represents the negative rule "silk tops cannot go with chiffon bottoms". Checking the rules respectively derived from the same metadata for AKD-DBPR and PKD-DBPR, we observed that PKD-DBPR outperforms AKD-DBPR especially when the given sample meets a weak matching rule that would be discarded by AKD but considered by PKD. For example, the first sample in Figure 5 activates the matching rule "the red top goes better with the white bottom than the pink one", which would be ignored by AKD but considered to be a weak rule for PKD as $fq(red, white)$ is slightly larger than $fq(red, pink)$ according to our dataset. In a sense, the ability of encoding more fuzzy human knowledge contributes to the better performance of PKD-DBPR. Unfortunately, PKD-DBPR can also yield several failed triplets, especially when the rules triggered by the given sample triplet

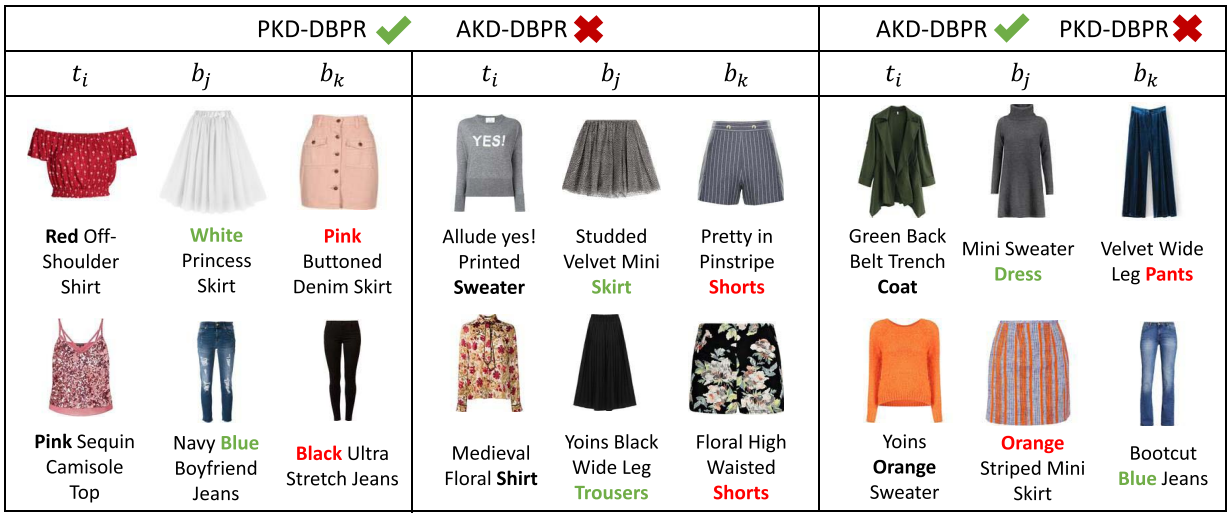
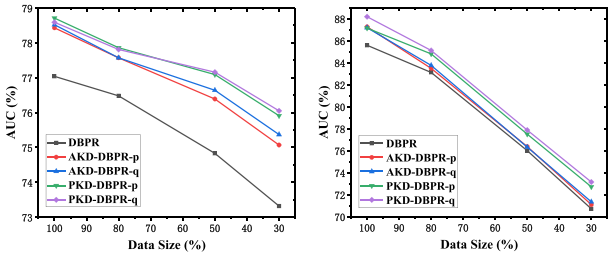


Fig. 5. Comparison between PKD-DBPR and AKD-DBPR on several testing triplets. All the triplets satisfy the ground truth that $t_i : b_j \succ b_k$, where "$\succ$" denotes the pairwise preference. We list the keywords of the metadata of items and bold the values of the rules. For each triplet, the green value refers to the higher co-occurrence frequency with the the black value of top t_i, compared with the red one.



(a) FashionVC (b) ExpFashion

Fig. 6. Comparison among different approaches with different dataset sizes.

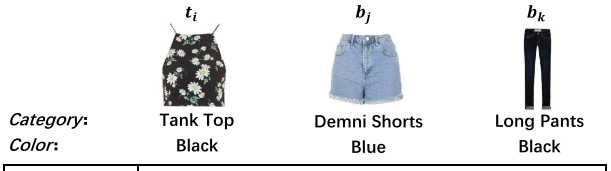
TABLE V
COMPARISON OF KNOWLEDGE ENCODING METHODS WITH RULES THEY COMMONLY OWNED IN TERMS OF AUC (%) ON TWO DATASETS

Approach	FashionVC	ExpFashion
AKD-DBPR-p	77.67	**86.85**
AKD-DBPR-q	**77.97**	86.77
PKD-DBPR-p	77.82	86.52
PKD-DBPR-q	77.71	86.52
AKD-PKD-DBPR-p	77.62	86.72
AKD-PKD-DBPR-q	77.79	86.64

are much comparable. For example, due to the fact that the coat can go with either the dress or pants, the matching rules activated by the first sample in the rightmost column would be encoded softly by PKD but hard by AKD, where the coats would be encouraged to go with the dress than pants. In this case that the given dress and pants both seem to be visually compatible with the coat, the hard rule guidance can be more powerful.

In addition, to evaluate the contribution of the fashion domain knowledge in reducing the data dependency, we also explored the performance of knowledge distillation methods with different amounts of training samples. Figure 6 illustrates the performance comparison of different approaches with different sizes of the training dataset. As we can see, the performance of all the approaches decreases gradually when the dataset size decreases from 100% to 30%, which is reasonable as the more training samples the model is fed, the better performance the model can achieve. Moreover, we found that AKD-DBPR and PKD-DBPR both outperform DBPR consistently at different sizes in both datasets, and the performance improvement grows with the number of training samples decreases, indicating the effectiveness of the fashion domain knowledge in scenarios with less labeled data samples. Besides, we also observed that the performance declining speed of PKD-DBPR is lower than AKD-DBPR which may be attributed again to the relatively abundant knowledge rules covered by PKD-DBPR.

Moreover, as both AKD-DBPR and PKD-DBPR are effective in knowledge distillation, to gain more thorough insights, we further integrated these two methods and obtained a new derivative AKD-PKD-DBPR, where the attentive rule reward of AKD-DBPR and probabilistic rule reward of PKD-DBPR are fused with equal weights in the knowledge distillation process. Notably, for fairness, we employed the common rules adopted by AKD-DBPR and PKD-DBPR. Table V shows the performance comparison among different methods. Interestingly, we found that the performance of all these methods are comparable. One possible explanation is that the effects of the probabilistic rule reward in PKD-DBPR and attentive rule reward in AKD-DBPR are essentially similar. To intuitively illustrate such similarity, we listed the attentive rule reward in AKD-DBPR and probabilistic rule reward in PKD-DBPR with a testing triplet example in Figure 7. As we can see, although AKD-DBPR and PKD-DBPR set different rule confidences and rule probabilities for different rules, respectively, the total rule rewards for the testing triplet of both methods are generally consistent. This reconfirms the advantage of PDK-DBPR in simplifying the rule confidence assignment to certain extent.



Rule for AKD	Rule 1: *tank top* + *shorts* (triggered by t_i and b_j) Rule 2: *black* + *black* (triggered by t_i and b_k)	
AKD-DBPR	Confidence of Rule 1	0.66
	Confidence of Rule 2	0.34
	Rule Reward on $[m^p_{ij}, m^p_{ik}]$	[0.66, 0.34]
Rule for PKD	Rule 1: $[fq(tank\ top, shorts), fq(tank\ top, pants)]$ Rule 2: $[fq(black, blue), fq(black, black)]$	
PKD-DBPR	Probability of Rule 1	[0.78, 0.22]
	Probability of Rule 2	[0.56, 0.44]
	Rule Reward on $[m^p_{ij}, m^p_{ik}]$	[0.67, 0.33]

Fig. 7. Illustration of the rule confidence setting.

TABLE VI
EFFECTS OF INTERNAL AND EXTERNAL RULES OF PKD-DBPR IN TERMS OF AUC (%) ON FASHIONVC

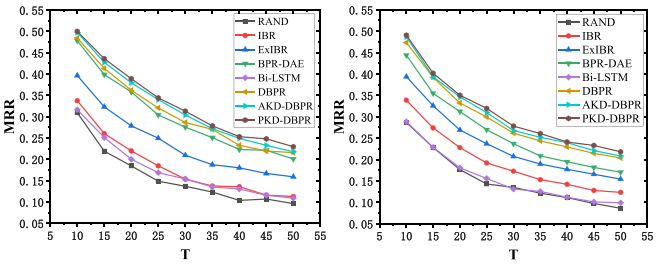
Approach	Internal Rule		External Rule	
	Category	Color	Category	Color
DBPR	77.04	77.04	77.04	77.04
PKD-DBPR-p	78.36	77.95	78.11	77.67
PKD-DBPR-q	78.24	77.98	78.24	78.14

D. Analysis on Rule Guidance

Furthermore, to comprehensively verify the rule guidance on PKD-DBPR, we evaluated PKD-DBPR on FashionVC with different knowledge rules, where we took into account both the internal and external rules. In particular, we treated the matching rules obtained from FashionVC as the internal rules, and those derived from all the outfits of ExpFashion as the external ones, which would be of high representativeness due to the considerable quantity of ExpFashion. Table VI exhibits the performance of the student network and teacher network of PKD-DBPR with different rule configurations across different sources. Similarly, each rule configuration here constrains the attributes (i.e., the category or color) of items that can be utilized for the knowledge derivation for PKD-DBPR. The first row refers to the performance of the baseline DBPR. As can be seen, both the internal and external rules can boost the performance of the compatibility modeling, which indicates the satisfactory generality of PKD-DBPR pertaining to not only the internal knowledge but also the external one. In addition, we noticed that the rules regarding the attribute *category* consistently yield the better performance than that on the attribute *color*, in both internal and external settings. This validates the fact that the category related rules are more widely recognized by the public and can provide the better guidance to the data-driven compatibility modeling neural network again.

E. Fashion Item Retrieval

To assess the practical value of the proposed compatibility modeling scheme, we conducted experiments in the context



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

Fig. 8. Performance of different approaches on complementary fashion item retrieval.

Fig. 9. Ranking result illustration of PKD-DBPR, AKD-DBPR, and DBPR. The bottoms highlighted in the red boxes are the positive ones.

of the complementary fashion item retrieval. Considering that it is time-consuming to rank all the bottoms for each top, we utilized the common strategy [26] that feeds each top t_i appeared in S_{test} as a query, and randomly selected T bottoms as ranking candidates, where there is only one positive bottom. We then fed the candidates into the trained neural networks to acquire their latent representations and calculated the compatibility score m_{ij} according to Eqn. (2), 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 [46], [47].

Due to the sparsity of the real-world dataset, we found that there are 1,262 tops, i.e., 64.59% of the 1,954 unique tops in the testing set, have not been observed in S_{train}. To comprehensively evaluate the proposed scheme, we compared it with different models using different type of testing tops: observed testing tops and unobserved ones. Figure 8 shows the performance comparison among different approaches on the complementary fashion item retrieval. We found that PKD-DBPR and AKD-DBPR outperform all the other baselines consistently at different numbers of bottom candidates, which demonstrates the advantage of incorporating the domain knowledge in the complementary fashion item retrieval. In addition, PKD-DBPR and AKD-DBPR achieve satisfactory performance with both observed and unobserved tops, indicating their capabilities of handling the cold start problem. Moreover, we observed that

PKD-DBPR outperforms AKD-DBPR in both scenarios, confirming the superiority of the probabilistic knowledge encoding manner in the real application.

We also listed the intuitive ranking results of PKD-DBPR, AKD-DBPR and DBPR for several testing tops in Figure 9. As we can see, as the first example activates the rule "tank top + shorts", AKD-DBPR and PKD-DBPR both bring the shorts in the candidate list forward. Moreover, since PKD incorporates more fuzzy human knowledge and integrates the weak rule "the red top goes better with the white bottom than the pink one", the order of the positive bottom gets further boosted to the first place, resulting in the better performance of PKD-DBPR than AKD-DBPR.

V. Conclusion and Future Work

In this work, we present a knowledge-guided compatibility modeling scheme to fulfil the clothing matching task, which is able to learn from not only the specific data samples but also the general knowledge rules. Considering that the human knowledge regarding clothing matching can be vast and fuzzy, we introduce an effective knowledge encoding method, PKD, which compiles the matching rules into the pure data-driven neural network in a probabilistic manner, making it possible to cope with the abundant domain knowledge without the manual screening. Extensive experiments conducted on two real-world datasets FashionVC and ExpFashion verify the portability of our model and demonstrate the advantages of integrating the domain knowledge in the context of clothing matching. Moreover, we find that both the internal and external rules can boost the performance of PKD-DBPR, validating the portability of our model. Interestingly, we also notice that knowledge rules regarding the category attribute are more powerful than those pertaining to other attributes (e.g., *color*) in guiding the compatibility modeling.

In this work, we mainly focus on taking the domain knowledge into consideration to tackle the problem of general clothing matching, but ignore the factor of user personal preferences in clothing matching. Therefore, in the future, we plan to explore the potential of the user context in complementary clothing matching.

References

[1] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," in *Proc. UAI*, Jun. 2009, pp. 452–461.

[2] X. Song, F. Feng, X. Han, X. Yang, W. Liu, and L. Nie, "Neural compatibility modeling with attentive knowledge distillation," in *Proc. ACM SIGIR*, Jul. 2018, pp. 5–14.

[3] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," Sep. 2014, *arXiv:1409.0473*. [Online]. Available: https://arxiv.org/abs/1409.0473

[4] D. J. Hu, R. Hall, and J. Attenberg, "Style in the long tail: Discovering unique interests with latent variable models in large scale social E-commerce," in *Proc. ACM SIGKDD*, Aug. 2014, pp. 1640–1649.

[5] S. Liu *et al.*, "Hi, magic closet, tell me what to wear!" in *Proc. ACM MM*, Oct. 2012, pp. 619–628.

[6] X. Gu, Y. Wong, P. Peng, L. Shou, G. Chen, and M. S. Kankanhalli, "Understanding fashion trends from street photos via neighbor-constrained embedding learning," in *Proc. ACM MM*, Oct. 2017, pp. 190–198.

[7] Y. Li, L. Cao, J. Zhu, and J. Luo, "Mining fashion outfit composition using an end-to-end deep learning approach on set data," *IEEE Trans. Multimedia*, vol. 19, no. 8, pp. 1946–1955, Aug. 2017.

[8] X. Han, X. Song, J. Yin, Y. Wang, and L. Nie, "Prototype-guided attribute-wise interpretable scheme for clothing matching," in *Proc. ACM SIGIR*, Jul. 2019, pp. 785–794.

[9] Y. Lin, P. Ren, Z. Chen, Z. Ren, J. Ma, and M. de Rijke, "Explainable outfit recommendation with joint outfit matching and comment generation," Jul. 2018, *arXiv:1806.08977*. [Online]. Available: https://arxiv.org/abs/1806.08977

[10] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. IEEE CVPR*, Jun. 2008, pp. 1–8.

[11] J. McAuley, C. Targett, Q. Shi, and A. Van Den Hengel, "Image-based recommendations on styles and substitutes," in *Proc. ACM SIGIR*, Aug. 2015, pp. 43–52.

[12] R. He and J. McAuley, "VBPR: Visual Bayesian personalized ranking from implicit feedback," in *Proc. AAAI*, Feb. 2016, pp. 144–150.

[13] A. Severyn and A. Moschitti, "Twitter sentiment analysis with deep convolutional neural networks," in *Proc. ACM SIGIR*, Aug. 2015, pp. 959–962.

[14] H. Jun, Q. Shengsheng, F. Quan, and X. Changsheng, "Attentive interactive convolutional matching for community question answering in social multimedia," in *Proc. ACM MM*, Oct. 2018, pp. 456–464.

[15] J. Liu, Y. Li, S. Song, J. Xing, C. Lan, and W. Zeng, "Multi-modality multi-task recurrent neural network for online action detection," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.

[16] Y. Zhang, J. Liu, W. Yang, and Z. Guo, "Image super-resolution based on structure-modulated sparse representation," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2797–2810, Sep. 2015.

[17] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *CoRR*, 2015.

[18] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," *Proc. IEEE*, vol. 105, no. 12, pp. 2295–2329, Dec. 2017.

[19] N. Papernot, P. McDaniel, X. Wu, S. Jha, and A. Swami, "Distillation as a defense to adversarial perturbations against deep neural networks," in *Proc. IEEE Symp. Secur. Privacy*, May 2016, pp. 582–597.

[20] R. Anil, G. Pereyra, A. Passos, R. Ormandi, G. E. Dahl, and G. E. Hinton, "Large scale distributed neural network training through online distillation," Apr. 2018, *arXiv:1804.03235*. [Online]. Available: https://arxiv.org/abs/1804.03235

[21] Z. Hu, X. Ma, Z. Liu, E. Hovy, and E. Xing, "Harnessing deep neural networks with logic rules," in *Proc. ACL*, 2016, pp. 1670–1679.

[22] R. Yu, A. Li, V. I. Morariu, and L. S. Davis, "Visual relationship detection with internal and external linguistic knowledge distillation," in *Proc. IEEE ICCV*, Oct. 2017, pp. 1974–1982.

[23] T. Alashkar, S. Jiang, S. Wang, and Y. Fu, "Examples-rules guided deep neural network for makeup recommendation," in *Proc. AAAI*, Feb. 2017, pp. 941–947.

[24] J. Rajendran, M. M. Khapra, S. Chandar, and B. Ravindran, "Bridge correlational neural networks for multilingual multimodal representation learning," Oct. 2015, *arXiv:1510.03519*. [Online]. Available: https://arxiv.org/abs/1510.03519

[25] D. Cao, L. Nie, X. He, X. Wei, S. Zhu, and T.-S. Chua, "Embedding factorization models for jointly recommending items and user generated lists," in *Proc. ACM SIGIR*, Aug. 2017, pp. 585–594.

[26] X. He, H. Zhang, M.-Y. Kan, and T.-S. Chua, "Fast matrix factorization for online recommendation with implicit feedback," in *Proc. ACM SIGIR*, Jul. 2016, pp. 549–558.

[27] L. Nie, X. Song, and T.-S. Chua, "Learning from multiple social networks," *Synthesis Lectures Inf. Concepts, Retr., Services*, vol. 8, no. 2, pp. 1–118, Apr. 2016.

[28] Z. Hu, Z. Yang, R. Salakhutdinov, and E. Xing, "Deep neural networks with massive learned knowledge," in *Proc. EMNLP*, Nov. 2016, pp. 1670–1679.

[29] Y. Ma, J. Jia, S. Zhou, J. Fu, Y. Liu, and Z. Tong, "Towards better understanding the clothing fashion styles: A multimodal deep learning approach," in *Proc. AAAI*, Feb. 2017, pp. 38–44.

[30] X. Song, F. Feng, J. Liu, Z. Li, L. Nie, and J. Ma, "Neurostylist: Neural compatibility modeling for clothing matching," in *Proc. ACM MM*, Oct. 2017, pp. 753–761.

[31] X. Song, L. Nie, L. Zhang, M. Akbari, and T.-S. Chua, "Multiple social network learning and its application in volunteerism tendency prediction," in *Proc. ACM SIGIR*, Aug. 2015, pp. 213–222.

[32] X. Song, L. Nie, L. Zhang, M. Liu, and T.-S. Chua, "Interest inference via structure-constrained multi-source multi-task learning," in *Proc. IJCAI*, Jun. 2015, pp. 2371–2377.
[33] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. EMNLP*, 2014, pp. 1746–1751.
[34] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. NIPS*, 2013, pp. 3111–3119.
[35] T. Sun, Y. Wang, J. Yang, and X. Hu, "Convolution neural networks with two pathways for image style recognition," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4102–4113, Sep. 2017.
[36] L. Song *et al.*, "A deep multi-modal CNN for multi-instance multi-label image classification," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 6025–6038, Dec. 2018.
[37] Y. You, C. Lu, W. Wang, and C.-K. Tang, "Relative CNN-RNN: Learning relative atmospheric visibility from images," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 45–55, Jan. 2019.
[38] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. ACM MM*, Nov. 2014, pp. 675–678.
[39] S. Rendle and L. Schmidt-Thieme, "Pairwise interaction tensor factorization for personalized tag recommendation," in *Proc. ACM WSDM*, Feb. 2010, pp. 81–90.
[40] H. Zhang, Z.-J. Zha, Y. Yang, S. Yan, Y. Gao, and T.-S. Chua, "Attribute-augmented semantic hierarchy: Towards bridging semantic gap and intention gap in image retrieval," in *Proc. ACM MM*, Oct. 2013, pp. 33–42.
[41] L. Bottou, "Stochastic gradient learning in neural networks," *Proc. Neuro-Nimes*, vol. 91, no. 8, p. 12, Nov. 1991.
[42] Z. Cheng, Y. Ding, L. Zhu, and M. Kankanhalli, "Aspect-aware latent factor model: Rating prediction with ratings and reviews," in *Proc. WWW*, Apr. 2018, pp. 639–648.
[43] G. Ding, Y. Guo, J. Zhou, and Y. Gao, "Large-scale cross-modality search via collective matrix factorization hashing," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5427–5440, Jun. 2016.
[44] J. Tang, K. Wang, and L. Shao, "Supervised matrix factorization hashing for cross-modal retrieval," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3157–3166, Jul. 2016.
[45] X. Han, Z. Wu, Y.-G. Jiang, and L. S. Davis, "Learning fashion compatibility with bidirectional LSTMs," in *Proc. ACM MM*, Oct. 2017, pp. 1078–1086.
[46] J. Yao, Y. Wang, Y. Zhang, J. Sun, and J. Zhou, "Joint latent Dirichlet allocation for social tags," *IEEE TMM*, vol. 20, no. 1, pp. 224–237, Jan. 2018.
[47] H. Zhang, Z. Kyaw, S.-F. Chang, and T.-S. Chua, "Visual translation embedding network for visual relation detection," in *Proc. CVPR*, Jul. 2017, vol. 1, no. 2, pp. 5532–5540.

Xianjing Han received the B.E. degree from Northeastern University, China, in 2017. She is currently pursuing the Ph.D. degree with the School of Computer Science and Technology, Shandong University, under the supervision of L. Nie and X. Song. Her research interests include multimedia computing and information retrieval.

Xuemeng Song received the B.E. degree from the University of Science and Technology of China, in 2012, and the Ph.D. degree from the School of Computing, National University of Singapore, in 2016. She is currently an Assistant Professor with Shandong University, Jinan, China. She has published several papers in the top venues, such as ACM SIGIR, MM, and TOIS. Her research interests include the information retrieval and social network analysis. In addition, she has served as a reviewer for many top conferences and journals.

Yiyang Yao received the master's degree in computer science from Northwestern Polytechnical University, Xi'an, China, in 2010. He is currently a member of State Grid Zhejiang Electric Power Company Ltd. His research interests include image processing and pattern recognition.

Xin-Shun Xu received the M.S. degree in computer science from Shandong University, China, in 2002, and the Ph.D. degree in computer science from Toyama University, Japan, in 2005. He is currently a Professor with the School of Software, Shandong University, where he joined the School of Computer Science and Technology, as an Associate Professor, in 2005. He joined the LAMDA Group, Nanjing University, China, as a Postdoctoral Fellow, in 2009. From 2010 to 2017, he was a Professor with the School of Computer Science and Technology, Shandong University. He is the Founder and the Leader of the Machine Intelligence and Media Analysis (MIMA) Group, Shandong University. His research interests include machine learning, information retrieval, data mining, and image/video analysis and retrieval. He has served as a Program Committee Member and a Reviewer for various international conferences and journals, including AAAI, IJCAI, MM, TIP, TKDE, and TMM.

Liqiang Nie received the B.Eng. degree from Xi'an Jiaotong University in July 2009 and the Ph.D. degree from the National University of Singapore (NUS), in 2013. He is currently a Professor with the School of Computer Science and Technology, Shandong University. He is the Adjunct Dean of the Shandong AI Institute. After the Ph.D. degree, he continued his research with NUS, as a Research Fellow for more than three years. He has coauthored over 140 papers. Meanwhile, he was supported by the Thousand Youth Talents Plan 2016 Program, Qilu Scholar 2016, and The Shandong Province Science Fund for Distinguished Young Scholars 2018. In 2017, he co-founded the Qilu Intelligent Media Forum. His research interests lie primarily in multimedia computing and information retrieval. He received over 4 700 Google Scholar citations as of June 2019. He is an AE of information science, an Area Chair of ACM MM 2018, a Special Session Chair of PCM 2018, and the PC Chair of ICIMCS 2017.