Visually-aware Recommendation with Aesthetic Features

Wenhu Yu, Xiangnan He, Jian Pei, Fellow, IEEE, Xu Chen, Li Xiong, Member, IEEE, Jinfei Liu, Member, IEEE, and Zheng Qin*

Abstract—Visual information plays a critical role in human decision-making process. While recent developments on visually-aware recommender systems have taken the product image into account, none of them has considered the aesthetic aspect. We argue that the aesthetic factor is very important in modeling and predicting users’ preferences, especially for some fashion-related domains like clothing and jewelry. This work addresses the need of modeling aesthetic information in visually-aware recommender systems. Technically speaking, we make three key contributions in leveraging deep aesthetic features: (1) To describe the aesthetics of products, we introduce the aesthetic features extracted from product images by a deep aesthetic network. We incorporate these features into recommender system to model users’ preferences in the aesthetic aspect. (2) Since in clothing recommendation, time is very important for users to make decision, we design a new tensor decomposition model for implicit feedback data. The aesthetic features are then injected to the basic tensor model to capture the temporal dynamics of aesthetic preferences (e.g., seasonal patterns). (3) We also use the aesthetic features to optimize the learning strategy on implicit feedback data. We enrich the pairwise training samples by considering the similarity among items in the visual space and graph space; the key idea is that a user may likely have similar perception on similar items. We perform extensive experiments on several real-world datasets and demonstrate the usefulness of aesthetic features and the effectiveness of our proposed methods.

Index Terms—Side information, aesthetic features, tensor factorization, pairwise learning to rank.

1 INTRODUCTION

RECOMMENDER systems have been widely used in online services to predict users’ preferences based on their interaction histories. Recently, visual information has been intensively explored to enhance the performance of recommender models. In many domains of interest, e.g., E-commerce and social media, the images of items play an important role in user decision-making process. For example, when purchasing clothing, users will scrutinize product images for the information like design, color schemes, decorative pattern, texture, and so on. To leverage these kinds of information, existing efforts have extracted various visual features from item images and injected them into recommender models, like SIFT features, CNN features, color histograms, etc. For example, utilized low-level SIFT features and color histograms, and utilized high-level CNN features extracted by a deep convolutional neural network. Despite these efforts, an important factor in visual information, aesthetics, has yet been considered in recommendation to the best of our knowledge.

*Z. Qin is corresponding author.

We argue that the aesthetic information is crucial and should not be ignored in predicting user preferences on products in many domains, such as clothing, furniture, food, electronics, etc. Taking the product shown in Figure 1 as an example, besides the semantic information, a user will also notice that the dress is with colors black and white, simple yet elegant design, and delightful proportion. She may have the intention to purchase it if she is satisfied with these aesthetic factors. In fact, for many users, especially young females, the aesthetic factor could be the primary factor — even more important than other common factors.
like quality, prices, and brand (see Figure 2). Unfortunately, conventional visual features do not encode the aesthetic information by nature. A recent work by Zhao et al. used color histograms to portray users’ intuitive perception about an image, but the solution leaves much space to improve, since it does not make good use of many other valuable information (such as aesthetic information shown in Figure 1). To address this issue, we need to leverage more comprehensive and high-level aesthetic features.

There are several ways to decompose a tensor [12], [13]. However, there are certain drawbacks in the existing models. To tailor it for the clothing recommendation task, we propose a new tensor factorization model trained with coupled matrices to mitigate the sparsity problem [14]. We then combine the basic model with the additional image features (concatenated aesthetic and CNN features) and term the method Visually-aware Recommendation with Aesthetic Features (VRA).

The other technical contribution of the paper lies in the learning part. When optimizing a model on implicit feedback data (e.g., purchasing records), pairwise learning has been widely used due to its rationality, which aims to maximize the margin between the predictions of positive and negative samples [15]. In this paper, we design a Multi-objective Faced Personalized Ranking (MPR) method to factorize the tensor and coupled matrices. However, when employing pairwise learning, one critical issue is that not all unobserved feedbacks are necessarily negative samples, since some of them might be just unknown by users. To address this issue, we leverage the visual information and collaborative information to construct the neighbor set of each item, and proposed a Neighbor-enhanced MPR (NMPR) algorithm. The intuition is that the items in the neighbor set of an observed item are less likely to be negative samples. Finally, we evaluate the performance of our proposed method by comparing it with several baselines on an Amazon dataset and 5 subsets. Extensive experiments show that the recommendation accuracy can be significantly improved by incorporating aesthetic features.

To summarize, our main contributions are as follows:

- We leverage aesthetic features to capture users’ aesthetic preferences in recommendation. Moreover, we compare the effectiveness with several conventional features to demonstrate the necessity of the aesthetic features.
- We propose a new tensor factorization model to portray the purchase events in three dimensions: users, items, and time. We then inject the aesthetic features into it to propose the hybrid VRA model and train it with coupled matrices to alleviate the sparsity problem.
- We propose a pairwise ranking method MPR for the multi-objective optimization. To enrich the pairwise training samples, we construct neighbor set for positive items by considering the similarity between items evidenced by visual features and collaborative information.

### 2 RELATED WORK

This paper develops aesthetic-aware clothing recommender systems. Specifically, we incorporate the features extracted from the product images by an aesthetic network into a tensor factorization model. As such, we review related work on aesthetic networks, image-based recommendation, tensor factorization, and enhanced pairwise learning.

#### 2.1 Aesthetic Networks

The aesthetic networks are proposed for image aesthetic assessment. After [16] first proposed the aesthetic assess-
ment problem, many research efforts exploited various handcrafted features to extract the aesthetic information of images [16], [17], [18]. To portray the subjective and complex aesthetic perception, [9], [10], [11] exploited deep networks to emulate the underlying complex neural mechanisms of human perception, and displayed the ability to describe image content from the primitive level (low-level) features to the abstract level (high-level) features. Proposed in [10], Brain-inspired Deep Network (BDN) model is the state-of-the-art aesthetic deep model. In this paper, we use BDN to extract the aesthetic features of product images, and use these features to enhance the performance of the recommender system.

2.2 Image-based Recommendations

Recommendation has been widely studied due to its extensive use, and many effective methods have been proposed [15], [19], [20], [21], [22], [23]. The power of recommender systems lies on their ability to model complex preferences that users exhibit toward items based on their past interactions and behavior. To extend their expressive power, various works exploited image data [3], [4], [5], [6]. For example, [4], [7] leveraged textual and visual information to recommend tweets and personalized key frames respectively; [3], [5], [6] used CNN features of product images while [5] recommended movies with color histograms of posters and frames. [24], [25], [26] recommended clothes by considering the clothing fashion style. Though various visual features are leveraged in recommendation tasks, they are conventional features (such as CNN features and SIFT features) and low-level incomplete aesthetic features (such as color histograms). In this paper, we propose high-level aesthetic features, which is extracted by a deep neural network, to take users’ aesthetic preference into account when recommend.

2.3 Tensor Factorization

Time is an important contextual information in recommender systems since the sales of commodities show a distinct time-related succession. In context-aware recom- mender systems, tensor factorization has been extensively used. For example, [12] introduced two main forms of tensor decomposition, the CANDECOMP/PARAFAC (CP) and Tucker decomposition. [27] first utilized tensor factorization for context-aware collaborative filtering. [13], [28] proposed a Pairwise Interaction Tensor Factorization (PITF) model to decompose the tensor with a linear complexity. In addition, tensor-based methods suffer from several drawbacks like poor convergence in sparse data [29] and not scalable to large-scale datasets [30]. To address these limitations, [14], [31], [32] formulated recommendation models with the Coupled Matrix and Tensor Factorization (CMTF) framework. All existing tensor decomposition models are designed for explicit feedback data and usually do not perform well in implicit feedback cases. In this paper, we design a novel tensor decomposition model for implicit feedback data and incorporate aesthetic features into it.

2.4 Enhanced Pairwise Learning

In real-world application, data of implicit feedback, or one-class form is easier to collect so extensively used. Prediction on implicit feedback dataset is a challenging work since we only know positive samples and unobserved samples, but cannot discriminate negative samples and potential positive samples from the unobserved ones [33]. In [15], all unobserved samples are treated equally as negative ones when sampling. To improve the sampling quality, many works proposed enhanced pairwise learning with various extra information [34], [35], [36], [37], [38]. For example, [34], [35] used view information to enrich positive samples. [36] proposed dynamic negative sampling strategies to maximize the utility of a gradient step by choosing “difficult” negative samples. [37], [38], [39] utilized collaborative information mined from the connection of users and items. [40], [41] proposed listwise ranking methods instead of pairwise ones. In this paper, we propose a visually-aware recommender model. Besides providing side information for prediction, the visual features are also used in the learning to rank process. For each positive sample, we regard items with similar visual features or items connected in the bipartite graph as the neighbors (potential positive samples), and assume that users will prefer them to other negative samples.

3 PRELIMINARIES

In this section, we introduce some preliminaries about the aesthetic neural network, which is used to extract the aesthetic features of clothing images. [10] introduced the Brain-inspired Deep Networks (BDN, shown in Figure 3), a deep CNN structure consists of several parallel pathways (sub-networks) and a high-level synthesis network. It is trained on the Aesthetic Visual Analysis (AVA) dataset, which contains 250,000 images with aesthetic ratings and tagged with 14 photographic styles (e.g., complementary colors, duotones, rule of thirds, etc.). The pathways take the form of convolutional networks to extract the abstracted aesthetic features by pre-trained with the individual labels of each tag. For example, when training the pathway for complementary colors, the individual label is 1 if the sample is tagged with “complementary colors” and is 0 if not. We input the raw features, which include low-level features (hue, saturation, value) and abstracted features (feature maps of the pathways), into the high-level synthesis network and jointly tune it with the pathways for aesthetic rating prediction. Considering that the AVA is a photography dataset and the styles are for photography, so not all the raw features extracted by the pathways are desired in our recommendation task, thus we only reserve the pathways that are relevant to the clothing aesthetic. Finally, we use the output of the second fully-connected layer of the synthesis network as our aesthetic features.

We then analyze several extensively used features to illustrate the superiority of our aesthetic features.

CNN Features: These are the most extensively used features due to their extraordinary representation ability. Typically the output of certain fully-connected layer of a deep CNN structure is used. For example, a common choice is the Caffe reference model with 5 convolutional layers followed by 3 fully-connected layers (pre-trained on
the ImageNet dataset); the features are the output of FC7, namely, the second fully-connected layer, which is a feature vector of length 4096.

CNN features mainly contain semantic information, which contributes little to evaluate the aesthetics of an image. Recall the example in Figure 1, it can encode “There is a skirt in the image” but cannot express “The clothing is beautiful and fits the user’s taste”. Devised for aesthetic assessment, BDN can capture the high-level aesthetic information. As such, our aesthetic features can do better in beauty estimating and complement CNN features in clothing recommendation.

Color Histograms: [5] exploited color histograms to represent human’s feeling about the posters and frames for movie recommendation. Though can get the aesthetic information roughly, the low-level handcrafted features are crude, unilateral, and empirical. BDN can get abundant visual features by the pathways. Also, it is data-driven, since the rules to extract features are learned from the data. Compared with the intuitive color histograms, our aesthetic features are more objective and comprehensive. Recall the example in Figure 1 again, color histograms can tell us no more than “The clothes in the image is white and black”.

4 AESTHETIC-BASED RECOMMENDATION

In this section, we first introduce the basic tensor factorization model, and then integrate image features into the basic model to propose the Visually-aware Recommendation with Aesthetic Features (VRA) model. The summary of notations are represented in Table 1.

4.1 Basic Model

Considering the impact of time on aesthetic preferences, we propose a context-aware model as the basic model to account for the temporal factor. We use a $P \times Q \times R$ tensor $A$ to indicate the purchase events among the user, clothes, and time dimensions (where $P$, $Q$, $R$ are the number of users, clothes, and time intervals, respectively). If user $p$ purchased item $q$ in time interval $r$, $A_{pqr} = 1$, otherwise $A_{pqr} = 0$. Tensor factorization has been widely used to predict the missing entries (i.e., zero elements) in $A$, which can be used for recommendation.

There are several approaches and we introduce the most common ones: Tucker Decomposition [12] has very strong representation ability, but it is very time consuming, and hard to converge. CP Decomposition [12], a simplification of Tucker Decomposition, has been widely used due to its linear time complexity [14], [31], [30], however, all dimensions are related by the same latent features thus the representation ability is weak. PITF Decomposition [13] is a balance of these two above methods, it has linear complexity and strong representation ability. Yet, it is not in line with implicit feedback applications due to the additive combination of each pair of matrices. For example, in PITF, for certain clothes $q$ liked by user $p$ but not fitting current time $r$, $q$ gets a high score for $p$ and a low score for $r$. Since we want to recommend the right item in the right time, $q$ should not be recommended to $p$. However, the total score can be high enough if $p$ likes $q$ so much that $q$’s score for $p$ is really high. In this case, $q$ will be returned even it does not fit the time. In addition, PITF model is inappropriate to be trained with coupled matrices.

To address the limitations of the aforementioned models, we propose a new tensor factorization method which is for implicit feedback with linear complexity. When a user makes a purchase decision on a clothing product, there are two primary factors: if the product fits the user’s preferences and if it fits the time. A clothing product fits a user’s...
preferences if the appearance is appealing, the style fits the user’s tastes, the quality is good, and the price is acceptable. And a clothing product fits the time if it is in-season and fashionable. For user $p$, clothing $q$, and time interval $r$, we use the scores $S_1$ and $S_2$ to indicate how the user likes the clothing and how the clothing fits the time respectively. $S_1 = 1$ when the user likes the clothing and $S_2 = 0$ otherwise. Similarly, $S_2 = 1$ if the clothing fits the time and $S_2 = 0$ otherwise. The user will buy the clothing only if $S_1 = 1$ and $S_2 = 1$, so $A_{pqr} = S_1 \& S_2$. To make the formula differentiable, we can approximately formulate it as $A_{pqr} = S_1 \cdot S_2$. We present $S_1$ and $S_2$ in the form of matrix factorization: $S_1 = U_p^T V_{sq}$, $S_2 = T_q^T W_{sr}$, where $U \in \mathbb{R}^{K_1 \times P}$, $V \in \mathbb{R}^{K_1 \times Q}$, $T \in \mathbb{R}^{K_2 \times R}$, and $W \in \mathbb{R}^{K_2 \times Q}$. The prediction is then given by:

$$A_{pqr} = \left( U_p^T V_{sq} \right) \left( T_q^T W_{sr} \right).$$

We can see that in Equation (1), the latent features relating users and clothes are independent with those relating clothes and time. Though the $K_1$-dimensional vector $V_{sq}$ and the $K_2$-dimensional vector $W_{sr}$ are all latent features of clothing $q$, $V_{sq}$ captures the information about users’ preferences intuitively whereas $W_{sr}$ captures the temporal information of the clothing. Compared with CP decomposition, our model is more expressive in capturing the underlying latent patterns in purchases. Compared with PITF, combining $S_1$ and $S_2$ with & (approximated by multiplication) is helpful to recommend right clothing in right time. Moreover, our model is efficient and easy to train compared with the Tucker decomposition.

Example 1. We give an example to illustrate how our basic model works. There are three items ($q_1$, $q_2$, and $q_3$) and two latent feature spaces (the user latent space and time latent space). The user latent space encodes the users’ preference and the time latent space encodes the temporal characteristics of items. In our basic model, we map users and items into user latent space by $U$ and $V$, and map time intervals and items into time latent space by $T$ and $W$. In this example, we aim to recommend clothes to a user $p$ who likes simple and elegant clothes in summer time $r$. For clothing $q_1$, we can see that it fits $p$’s preference and it is a shirt designed for summer, thus $q_1$ gets high $S_1$ and $S_2$ scores and can be recommended due to the high score $S = S_1 \cdot S_2$. For the clothing $q_2$, it is a piece of summer clothes yet is too colorful for $p$, thus $q_2$ gets low $S_1$ score and high $S_2$ score and cannot be recommended. Clothing $q_3$ is simple and elegant yet is used in winter, thus $q_3$ gets high $S_1$ score and low $S_2$ score and cannot be recommended either.

4.2 Hybrid Model

In this section, we incorporate the visual features into the basic model, and optimize it with the pairwise learning to rank method.

4.2.1 Problem Formulation

Combined with image features, we formulate the predictive model as:

$$A_{pqr} = \left( U_p^T V_{sq} + M_{sq}^T F_{sq} \right) \left( T_q^T W_{sr} + N_{sr}^T F_{sr} \right),$$

where $F \in \mathbb{R}^{K \times Q}$ is the feature matrix, $F_{sq}$ is the image features of clothing $q$, which is the concatenation of CNN features ($f_{CNN}$) and aesthetic features ($f_{AES}$), $F_{sr} = \begin{bmatrix} f_{CNN} \\ f_{AES} \end{bmatrix}$ and $K = 8192$. $M \in \mathbb{R}^{K \times P}$ and $N \in \mathbb{R}^{K \times R}$ are aesthetic preference matrices. $M_{sq}$ encodes the preferences of user $p$ and $N_{sr}$ encodes the preferences in time interval $r$. In our model, both the latent features and image features contribute to the final prediction. Though the latent features can uncover any relevant attributes theoretically, they usually cannot in real-world applications on account of the sparsity of the data and lack of information. So the assistance of image information can highly enhance the model. Also, recommender systems often suffer from the cold start problem. We cannot extract information for users and clothes without consumption records. In this case, content and context information can alleviate this problem. For example, for certain “cold” clothing $q$, we can decide whether to recommend it to certain user $p$ in current time $r$ according to if $q$ looks satisfying to the user (determined by $M_{sp}$) and to the time (determined by $N_{sr}$).
4.2.2 Coupled Matrix and Tensor Factorization

Though widely used to portray the context information in recommendation, tensor factorization suffers from poor convergence due to the sparsity of the tensor. To relieve this problem, [14] proposed a CMTF model, which decomposes the tensor with coupled matrices. In this subsection, we couple our tensor factorization model with restrained matrices during training.

User × Clothing Matrix: We use matrix $B \in \mathbb{R}^{P \times Q}$ to indicate the purchase activities between users and clothes. $B_{pq} = 1$ if user $p$ purchased clothing $q$ and $B_{pq} = 0$ if not.

Time × Clothing Matrix: We use matrix $C \in \mathbb{R}^{R \times Q}$ to record when the clothing was purchased. Since the characteristics of clothing change steadily with time, we make a coarse-grained discretization on time to avoid the tensor from being extremely sparse. Time is divided into $R$ intervals in total. $C_{rq} = 1$ if clothing $q$ is purchased in time interval $r$ and $C_{rq} = 0$ if not.

4.2.3 Multi-objective Faced Personalized Ranking

In this subsection, we design pairwise learning method Multi-objective Faced Personalized Ranking (MPR) for our VRA. We represent the positive set $D$ in the form of triples:

$$D = \{(p, q, r) | \hat{A}_{pqr} = 1\},$$

and the set of unlabeled samples is:

$$Q_{pr} = \{q | q \in Q \setminus (Q_p^+ \cup Q_r^+)\},$$

where $Q$ denotes the set of items, $Q_p^+ = \{q | B_{pq} = 1\}$ denotes the set of items purchased by user $p$, and $Q_r^+ = \{q | C_{rq} = 1\}$ denotes the set of items purchased in time $r$.

The objective function is formulated as:

$$\text{MPR}_{OPT} = \sum_{(p, q, r) \in D} \sum_{q' \in Q_{pr}} L(p, q, q', r) - \frac{\lambda_r}{2} ||\theta||^2_{F}. \quad (3)$$

$L(\cdot)$ in Equation (3) is the likelihood function,

$$L(p, q, q', r) = \ln \sigma(\hat{A}_{pqqr'}) + \lambda_c \left[ \ln \sigma(\mathbf{B}_{pqq'}) + \ln \sigma(\mathbf{C}_{rqq'}) \right],$$

where $\hat{A}$ is defined in the Equation (2), $\mathbf{B} = \mathbf{U}^T \mathbf{V} + \mathbf{M}^T \mathbf{F}$, and $\mathbf{C} = \mathbf{T}^T \mathbf{W} + \mathbf{N}^T \mathbf{F}$. $\hat{A}_{pqqr'} = \mathbf{A}_{pqq'r} - \mathbf{A}_{pqqr}, \mathbf{B}_{pqq'} = \mathbf{B}_{pqq'} - \mathbf{B}_{pqq'}, \mathbf{C}_{rqq'} = \mathbf{C}_{rqq'} - \mathbf{C}_{rq}, (\cdot)$ is the sigmoid function; $\lambda_c$ is a parameter to balance the weights of the tensor term and coupled matrix terms. The last term of Equation (3) is the regularization term to prevent overfitting, and $\lambda_r$ is the regularization coefficient. $||\theta||_F$ is the Frobenius norm of the matrix, $\Theta$ represents the parameters of the model, $\Theta = \{\mathbf{U}, \mathbf{V}, \mathbf{T}, \mathbf{W}, \mathbf{M}, \mathbf{N}\}$. The model is optimized from users’ implicit feedback with mini-batch gradient descent, which calculates the gradient with a small batch of samples.

5 Neighbor-enhanced Pairwise Learning

In the previous subsection, we introduced MPR, which is a pairwise learning method for multi-objective optimization, with the aim of maximizing the gap between the positive feedbacks and negative feedbacks. Pairwise learning has been widely used due to its strong performance [6, 8, 12] while there is a critical issue in the current formulation. To be specific, a user did not purchase a product may because she is not interested in it, but may also because that she has never seen it before. Our task is to predict the preferences of users and recommend them unseen products they are interested in. However, in pairwise learning, all missing entries are treated as negative samples. To address this gap, we construct the neighbor set $N_q$ for each positive sample $q$ by uncovering the products that have similar visual features with $q$, or the products connected to $q$ in the user-item or time-item graphs. In other words, $N_q$ contains the products near $q$ in the visual space or in the graph. In this section, we propose a Neighbor-enhanced MPR (NMPR) for our time-aware model with side information.

5.1 Problem Formulation

When sampling, we regard the neighbors as potential positive samples. For a user $p$ and a time interval $r$, we assume that (1) user $p$ prefers items with positive feedbacks to the others; (2) user $p$ prefers the neighbors of the positive sample to the irrelevant ones; (3) positive samples fit the current time $r$ better than the others; (4) neighbors of the positive sample fit the current time $r$ better than the irrelevant ones. So for each $(p, q, r)$ in $D$, we have the preference relationship,

$$(p, q, r) \succ (p, Q_{pr}, r), \quad (p, q, r) \succ (p, N_q, r), \quad (p, N_q, r) \succ (p, Q_{pr} \setminus N_q, r).$$

As such, we can generalize Equation (3) as follows:

$$\text{NMPR}_{OPT} = \sum_{(p, q, r) \in D} \left[ \sum_{q'' \in Q_{pr}} L(p, q, q'', r) + \eta_1 \sum_{q' \in N_q} L(p, q, q', r) + \eta_2 \sum_{q'' \in Q_{pr} \setminus N_q} L(p, q, q'', r) \right] - \frac{\lambda_r}{2} ||\theta||^2_{F}. \quad (4)$$

Here we can see that for each purchase record $(p, q, r)$, user $p$ prefers $q$ to $q'$ and prefers $q'$ to $q''$. The preference relationship is constructed by finding the neighbors of the positive items, which can be interpreted as an item collaborative learning model [23]. Most existing works learn to rank by constructing the potential set of each user [39, 37, 43, 34, 38, 55]. Now, we give an example to illustrate the advantage of our item collaborative learning model.

5.2 Constructing Neighbor Set

To find the neighbors of each positive sample, we leverage the visual information and the collaborative information. For visual information, we cluster all products with CNN...
features and aesthetic features. For each product, the cluster it belongs to is the neighborhood set. And for collaborative information, we find all products purchased by the same user or purchased in the same time to be the neighbor products.

**Neighbors in semantic space:** We cluster all products by the CNN features. For a product \( q \), the cluster it belongs to is the semantic neighbor set, denoted as \( N^C_q \). Products with similar CNN features have similar appearances, users may have interests in the items that look like the purchased ones.

**Neighbors in aesthetic space:** Similarly, we cluster all products by the aesthetic features and regard the cluster a product \( q \) belongs to as the aesthetic neighbor set, denoted as \( N^A_q \). Products close to each other in the aesthetic space have similar aesthetic characteristics, users may prefer the items which are in line with their aesthetics.

**Neighbors linked by users:** For each product \( q \), we find all products that purchased by the same user to construct the user-linked neighbor set, \( N^U_q \). Each product \( q' \) in \( N^U_q \) has been purchased by certain user with \( q \), users who have interests in \( q \) may also like \( q' \). We update the part of our model which captures the users’ preferences (parameters \( U, V \), and \( M \)) with \( N^V_q \).

**Neighbors linked by time:** For each product \( q \), we find all products that purchased in the same time with \( q \) to construct the time-linked neighbor set, \( N^T_q \). Each product \( q' \) in \( N^T_q \) has been purchased in the same time with the current product \( q \), so \( q' \) may fit the current time better than other missing value samples. We update the part which captures the temporal character of products in our model (parameters \( T, W \), and \( N \)) with \( N^T_q \).

### 5.3 Model Learning

We then calculate the gradient of Equation 4. To maximize the objective function, we take the first-order derivatives with respect to each model parameter:

\[
\nabla_{\Theta} \text{NMPR}_{op_t} = \sum_{(p,q,r) \in D} \left[ \sum_{q' \in \mathcal{N}^C_q} \frac{\partial L(p,q,q',r)}{\partial \Theta} + \eta_1 \sum_{q' \in \mathcal{N}^U_q} \frac{\partial L(p,q,q',r)}{\partial \Theta} + \eta_2 \sum_{q' \in \mathcal{N}^T_q} \frac{\partial L(p,q,q',r)}{\partial \Theta} \right] - \lambda_\epsilon \Theta.
\]

where

\[
\frac{\partial L(p,q,q',r)}{\partial \Theta} = \begin{cases} \sigma(-\hat{A}_{pq}q') \frac{\partial \hat{A}_{pq}q'}{\partial \Theta} \\ + \lambda_\epsilon \left[ \sigma\left(-\hat{B}_{pq}q'\right) \frac{\partial \hat{B}_{pq}q'}{\partial \Theta} + \sigma\left(-\hat{C}_{rq}q'\right) \frac{\partial \hat{C}_{rq}q'}{\partial \Theta} \right] \end{cases}.
\]

We use \( \Theta \) to denote certain column of \( \Theta \). For our VRA model, the derivatives are:

\[
\frac{\partial \hat{A}_{pq}q'}{\partial \Theta} = \begin{cases} \hat{C}_{rq}V_{s_\epsilon} - \hat{C}_{rq}V_{s'_\epsilon} & \text{if } \Theta = U_{s_\epsilon} \\ \hat{C}_{rq}V_{s_\epsilon} - \hat{C}_{rq}V_{s'_\epsilon} & \text{if } \Theta = V_{s_\epsilon} \\ \hat{C}_{rq}F_{s_\epsilon} - \hat{C}_{rq}F_{s'_\epsilon} & \text{if } \Theta = M_{s_\epsilon} \end{cases}
\]

\[
\frac{\partial \hat{B}_{pq}q'}{\partial \Theta} = \begin{cases} V_{s_\epsilon} - V_{s'_\epsilon} & \text{if } \Theta = U_{s_\epsilon} \\ U_{s_\epsilon} - U_{s'_\epsilon} & \text{if } \Theta = V_{s_\epsilon} \\ F_{s_\epsilon} - F_{s'_\epsilon} & \text{if } \Theta = M_{s_\epsilon} \end{cases}
\]

Equations (6) and (7) give the derivatives for \( \Theta = \{U, V, M\} \), and we can get the similar form for \( \Theta = \{T, W, N\} \). The partial derivative of \( L \) in Equation (6) is certain column of \( \frac{\partial \hat{A}_{pq}q'}{\partial \Theta} \) in Equation (5), for example, the \( p \)-th column when \( \Theta = U_{s_\epsilon} \).

Finally, we update the parameters with the derivatives we get. As discussed in Subsection 5.2, we use different neighborhood sets to update different parts of the model. For \( \Theta = \{U, V, M\} \), we update the parameters:

\[
\Theta = \Theta + \eta \nabla_{\Theta} \text{NMPR}_{op_t} \bigg|_{\mathcal{N}^U_q = \mathcal{N}^C_q = \mathcal{N}^A_q = \mathcal{N}^T_q = \emptyset}
\]

and for \( \Theta = \{T, W, N\} \):

\[
\Theta = \Theta + \eta \nabla_{\Theta} \text{NMPR}_{op_t} \bigg|_{\mathcal{N}^C_q = \mathcal{N}^A_q = \mathcal{N}^T_q = \emptyset}
\]

Our model is optimized with mini-batch gradient descent and for each positive sample, we sample \( \rho \) negative samples and \( \rho \) neighbors randomly to construct pairs, where \( \rho \) is the sampling rate.

### 6 Experiments

In this section, we conduct experiments on real-world datasets to verify the effectiveness of our method. We then analyze the experimental results and demonstrate the improvements over competing baselines. We focus on answering the following four key research questions:

**RQ1:** What factors affect users’ aesthetics?

**RQ2:** How is the performance of our overall solution for the clothing recommendation task?

**RQ3:** What are the advantages of the aesthetic features compared with conventional image features?

**RQ4:** How is the performance of our NMPR enhanced with collaborative and visual information?

#### 6.1 Experimental Setup

**6.1.1 Datasets**

We use the AVA dataset to train the aesthetic network and use the Amazon dataset to train the recommendation models.

- **Amazon:** The Amazon dataset [3] is the consumption records from Amazon.com. In this paper, we use the clothing shoes and jewelry category filtered with 5-core (remove users and items with less than 5 purchase records) to train all recommendation models. There are 39,371 users, 23,022 items, and 275,539 records in total (after 2010). The sparsity of the dataset is 99.969%.

- **Aesthetic Visual Analysis (AVA):** We train the aesthetic network with the AVA dataset [14], which is the collection of images and meta-data derived from DPCChallenge.com. It contains over 250,000 images with aesthetic ratings from 1 to 10, 66 textual tags describing the semantics of images, and 14 photographic styles (complementary colors, duotones, high dynamic range, image grain, light on white, long exposure, macro, motion blur, negative image, rule of thirds, shallow DOF, silhouettes, soft focus, and vanishing point).
6.1.2 Experiment Settings
In the Amazon dataset, we remove the record before 2010. Time is discretized by weeks, and there are 237 time intervals in total. To validate the scalability of the model and give a comprehensive assessment, we split the dataset into several subsets by gender and categories of products (jewelry dataset includes both jewelries and watches).

| Dataset  | Purchase | User | Item | Sparsity of Matrices/Tensors |
|----------|----------|------|------|-----------------------------|
| Amazon   | 275539   | 39371| 23022| 99.9696% / 99.9999%         |
| Men      | 67156    | 22547| 5460 | 99.9454% / 99.9998%        |
| Women    | 176136   | 35059| 14500| 99.9653% / 99.9999%       |
| Clothes  | 115841   | 32728| 8777 | 99.9597% / 99.9998%       |
| Shoes    | 94560    | 32538| 8231 | 99.9647% / 99.9999%       |
| Jewelry  | 37314    | 15924| 3607 | 99.9350% / 99.9997%       |

We then randomly split each dataset into training (80%), validation (10%), and test (10%) sets, and remove the cold items and users from the validation and test sets. The validation set is used for tuning hyper-parameters and the final performance comparison is conducted on the test set. We make the prediction and recommend the top-$n$ items to each user. The F1-score and the normalized discounted cumulative gain (NDCG) are calculated to evaluate the performance of the baselines and our model. Our experiments are conducted by predicting Top-5, 10, 20, 50, and 100 favourite clothing.

6.2 Influential Factors of Aesthetics (RQ1)
In this subsection, we explore some factors that impact the users’ aesthetics. HSV (Hue, Saturation, and Value), inputted as low-level aesthetics features in the BDN, are studied to show how aesthetic preferences change with the influence of certain factor.

![Fig. 6. Distribution of hue, saturation, and value of the whole dataset.](image)

Figure 6 shows the distribution of hue, saturation, and value, which are counted from the whole Amazon dataset. We normalize hue, saturation, and value into $[0, 1]$ and normalize the histograms into a unit vector. The bar in the bottom of Figure 6(a) is the hue, and different hue indicates different colors. From the figure we can see that users prefer red and blue. The bar in the bottom of Figure 6(b) is the saturation, which defines the brilliance and intensity of a color. From Figure 6(b) we can see that users prefer a lower saturation. The bar in the bottom of Figure 6(c) is the value, which refers to the lightness or darkness of a color. The larger the value is, the lighter the color is. To present the difference of aesthetic preferences with certain factor, we report the difference between the normalized HSV histograms before and after the influence of certain factor, so there are positive values and negative values (see Figures 7 to 10). We mainly discuss the variation of HSV with different kinds of users and in different time.

6.2.1 Influence of users
Modern recommender systems aim to provide the personalized recommendation, so the influence of different kinds of users is very important. It is obvious that different users have different aesthetic preferences. In this subsection, we show the variation of HSV impacted with the gender and age.

**Users with different ages:** Figure 7 shows the impact of users with different ages. Figure 7(a) and 7(b) show the saturation distribution of kids and adults, respectively. Kids like clothes with really high saturation while adults like those with low saturation.

![Fig. 7. Aesthetic preferences of users with different ages.](image)

**Users with different genders:** Figure 8 presents the aesthetic preferences of males and females. Figure 8(a) shows the distribution of the value with males. They prefer dark clothes that can make them look mature and steady. Figure 8(b) shows the distribution with females. They prefer lovely and active clothes in light colors.

![Fig. 8. Aesthetic preferences of users with different genders.](image)

6.2.2 Influence of time
For many products, especially clothes, movies, electronic devices, etc., sales change dramatically with time. Users’ aesthetic preferences also change with time. For example, people like different colors and design in different seasons. Also, the fashion changes every year. In this subsection, we represent how time influences aesthetic preferences in a short term and long term.

**Seasonality:** Figure 9 represents users’ aesthetic preferences in different seasons. Figures 9(a) to 9(d) show the distribution of value in spring, summer, autumn and winter, respectively. Users prefer light colors in spring and summer while prefer dark colors in autumn and winter.

**Annual trend:** The fashion trend in different years is shown in Figure 10. Histograms in Figures 10(a) to 10(c)
Fig. 9. Aesthetic preferences in different seasons.

show the hue distribution of clothes in 2010, 2012 and 2014, respectively. As shown in Figure 10, users preferred yellow and blue in 2010. In 2012, yellow and purple became popular. In 2014, the most popular color was red.

Fig. 10. Aesthetic preferences in different years.

From the figures above, we come to the conclusion that users’ aesthetic preferences change with different people and different time. So we propose a time-aware model taking these two factors into account as the basic model.

6.3 Performance of Our Model (RQ2)

To demonstrate the effectiveness of our model, we adopt the following methods as baselines for performance comparison:

- **Most Popular (MP):** This baseline ranks items according to their popularity thus is non-personalized.
- **PMF:** This Probabilistic Matrix Factorization method was proposed in [19], which is a frequently used state-of-the-art approach for rating-based optimization and prediction. We set the score of positive samples as 1 and missing values as 0.
- **BPR:** This Bayesian Personalized Ranking method is a well known ranking-based method [15] for implicit feedback. The preference pairs are constructed between the positive samples and the other ones. In our experiments, we randomly sample five negative instances for each positive feedback.
- **CMTF:** This Coupled Matrix and Tensor Factorization model is a state-of-the-art context-aware recommendation method [14]. The tensor factorization is jointly learned with several coupled matrices.
- **CPLR:** This Collaborative Pairwise Learning to Rank method [39] is an extension of BPR, which tries to relax BPR’s assumptions using the idea of collaborative filtering.
- **VBPR:** This Visual Bayesian Personalized Ranking method is a state-of-the-art visually-aware recommendation method [3]. The image features are pre-generated from the product images using the Caffe deep learning framework.

We iterate 200 times to train all models. The sampling rate $\rho$ is set as 5. In each iteration, we enumerate all positive records and select 1000 users in test/validation set to calculate evaluation metrics, and then record the best performance (for MP, we test 200 times without training). We show the $F_1$-score and NDCG with different $n$ in Figures 11 and 12 respectively. Subfigures (a) to (f) show the performance on Amazon, Men, Women, Clothes, Shoes and Jewelry, respectively.

For all datasets and all models, we repeat our experiments 10 times. The bars in Figures 11 and 12 indicate the average performance and the vertical lines on the top of the bars indicate the standard deviation. We can see that the datasets with higher sparsity show lower performance.

Compared with MP, personalized methods show stronger abilities to represent the preferences of users and outperform MP several times. By recommending clothes that fit the current season, CMTF outperforms PMF on both $F_1$-score and NDCG. Enhanced with side information, VBPR performs the best among all baselines. With the aesthetic features providing more information, VRA outperforms all baselines on all datasets. Taking Jewelry
as an example, the proposed VRA model outperforms VBPR about 13.38% on $F_1$-score@5 and 12.70% on NDCG@5. We also conduct $t$-tests, verifying that all improvements over VBPR are statistically significant for $p < 0.05$.

In our experiments, we tune the hyperparameters ($\lambda_c$ and $\lambda_r$) for all models on the validation set, the sensitivity analysis is shown in Figure 13 (take Jewelry as an example). $\lambda_c$ is a weighting parameter for the coupled matrices, the performance with different $\lambda_c$ is shown in Figure 13(a). Only VRA and CMTF are impacted with it. When $\lambda_c = 0.01$, our model achieves the best performance. The sensitivity with regularization coefficient is shown in Figure 13(b). We can see that with the variation of $\lambda_r$, point-wise optimized models (PMF and CMTF) response quite differently with pair-wise optimized ones (BPR, CPLR, VBPR, and VRA). When $\lambda_r$ is larger than 0.4, the $F_1$-score of PMF and CMTF reduces rapidly with the increasing of $\lambda_r$ while the $F_1$-score of BPR, CPLR, VBPR, and VRA still increase. The performance of all models are quite similar to each other before $\lambda_r = 0.4$. For pair-wise optimized models, the performance show different gradually after $\lambda_r = 0.8$. When $\lambda_r = 1.5$, our model performs the best.

6.4 Necessity of the aesthetic features (RQ3)

In this subsection, we discuss the necessity of the aesthetic features. We combine various widely used features to our basic model and compare the effectiveness of each feature by constructing five models:

- **VRA basic**: This is our basic Visually-aware Recommendation model without any image features, which is represented in Subsection 4.1.
- **VRH**: This is a Visually-aware Recommendation with Color Histograms.
- **VRCo**: This is a Visually-aware Recommendation with CNN Features only.
- **VRAo**: This is a Visually-aware Recommendation with Aesthetics Features only.
- **VRA**: This is our proposed model, utilizing both CNN features and aesthetic features.

All models are optimized on Jewelry dataset, we repeat the experiments 5 times and report the $F_1$-score and the NDCG.
Fig. 16. Items purchased by users and recommended by different models (Amazon dataset).

NDCG in Figures 15(a) and 15(b) respectively. As shown in Figure 15, VRA_basic performs the worst since no image features are involved to provide the extra information. With the information of color distribution, VRH performs better, though still worse than VRCo and VRAo, because the low-level features are too crude and unilateral, and can provide very limited information about users’ aesthetic preferences. VRCo and VRAo show the similar performance because both CNN features and aesthetic features have strong ability to mine user’s preferences. Our VRA model, capturing both semantic information and aesthetic information, performs the best on the dataset since those two kinds of information mutually enhance each other to a certain extent. Give an intuitive example, if a user wants to purchase a skirt, she needs to tell whether there is a skirt in the image (semantic information) when looking through products, and then she needs to evaluate if the skirt is good-looking and fits her tastes (aesthetic information) to make the final decision. In our experiments, VRA outperforms VRCo and VRAo about 6.42% and 9.08% on $F_1$-score@5, 6.03% and 8.58% on NDCG@5 respectively. We can see that though the aesthetic features and CNN features do not perform the best separately, they mutually enhance each other and achieve improvement together.

Several purchased and recommended items on Amazon dataset are represented in Figure 16. The items in the first row are purchased by certain user (training data, the number is random). To illustrate the effectiveness of the aesthetic features intuitively, we choose the users with explicit style of preferences and single category of items. The items in the second row and third row are recommended by VRCo and VRA respectively. For these two rows, we choose the five best items from the 50 recommendations to exhibit. Comparing the first and the second row, we can see that leveraging semantic information, VRCo can recommend the congeneric (with the CNN features) and relevant (with tensor factorization) commodities. Although it can recommend the pertinent products, they are usually not in the same style with what the user has purchased. Capturing both aesthetic and semantic information, VRA performs much better. We can see that the items in the third row have more similar style with the training samples than the items in the second row.

Taking Figure 16(c) as an example, we can see that the user prefers boots, ankle boots, or thigh boots. However, products recommended by VRCo are some different types of women’s shoes, like high heels, snow boots, thigh boots, and cotton slippers. Though there is a thigh boot, it is not in line with the user’s aesthetics due to the gaudy patterns and stumpy proportion, which rarely appears in her choices. The products recommended by VRA are better. First, almost all recommendations are boots. Then, thigh boots in the third row are in the same style with the training samples, like leather texture, slender proportions, simple design and some design elements of detail like straps and buckles (the second and third ones). Though the last one seems a bit different with the training samples, it is in the uniform style with them intuitively, since they are all designed for young ladies. It is also obvious in Figure 16(f) we can see that what the user likes are vibrant watches for young men. However, watches in the second row are in pretty different styles, like digital watches for children, luxuriantly-decorated ones for ladies, old-fashioned ones for adults. Evidently, watches in the third row are in similar style with the train samples. They have similar color schemes and design elements, like the intricated-designed dials, nonmetallic watchbands, small dials, and tachymeters. As we can see, with the aesthetic features and the CNN features complementing each other, VRA performs much better than VRCo.
6.5 Performance of NMPR (RQ4)

In this subsection, we illustrate the effectiveness of our NMPR optimization criterion.

Figure 17 shows the performance with different weighting parameters $\eta_1$ and $\eta_2$. We can see that when $\eta_1 = 0.1$ and $\eta_2 = 0.01$, the model achieves the best performance. When $\eta_1 = 0$ and $\eta_2 = 0$, the model becomes VRA-MPR. As shown in the figure, VRA-NMPR outperforms VRA-MPR about 4.70% in $F_1$-score.

When $\eta_2$ is fixed, $F_1$-score usually takes the maximum when $\eta_1$ is about 0.1. When $\eta_1$ is fixed, $F_1$-score usually takes the maximum when $\eta_2$ is about 0.01. We come to the conclusion that the preference relation $(p, q, r) \succ (p, N_q, r)$ is more important than $(p, N_q, r) \succ (p, N_p \setminus N_q, r)$.

7 CONCLUSION

In this paper, we investigated the usefulness of aesthetic features for personalized recommendation on implicit feedback datasets. We proposed a novel model that incorporates aesthetic features into a tensor factorization model to capture the aesthetic preferences of users at a particular time, and leveraged visual information and collaborative information to optimize the model. Experiments on challenging real-world datasets show that our proposed method dramatically outperforms state-of-the-art models, and succeeds in recommending items that fit users’ style.

For the future work, we are interested in constructing high-order connections among items with spectrum clustering, social networks, etc. instead of only one-order connections, to enhance the pairwise learning. Also, we will establish a large dataset for product aesthetic assessment, and train the networks to extract the aesthetic information better. Lastly, we will investigate the effectiveness of the aesthetic features in the setting of explicit feedback.

REFERENCES

[1] W. Yu, H. Zhang, X. He, X. Chen, L. Xiong, and Z. Qin, “Aesthetic-based clothing recommendation,” in WWW, 2018, pp. 649–658.
[2] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, “Neural collaborative filtering,” in WWW, 2017, pp. 173–182.
[3] R. He and J. McAuley, “VBPR: visual bayesian personalized ranking from implicit feedback,” in AAAI, 2016, pp. 144–150.
[4] X. Chen, Y. Zhang, Q. Ai, H. Xu, J. Yan, and Z. Qin, “Personalized key frame recommendation,” in SIGIR, 2017, pp. 315–324.
[5] L. Zhao, Z. Lu, S. J. Pan, and Q. Yang, “Matrix factorization+ for movie recommendation,” in IJCAI, 2016, pp. 3945–3951.
[6] J. Chen, H. Zhang, X. He, L. Nie, W. Liu, and T.-S. Chua, “Attentive collaborative filtering: Multimedia recommendation with item- and component-level attention,” in SIGIR, 2017, pp. 335–344.
[7] T. Chen, X. He, and M.-Y. Kan, “Context-aware image tweet modelling and recommendation,” in MM, 2016, pp. 1018–1027.
[8] J. McAuley, C. Taraffet, Q. Shi, and A. van den Hengel, “Image-based recommendations on styles and substitutes,” in SIGIR, 2015, pp. 43–52.
[9] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, “RAPID: rating pictorial aesthetics using deep learning,” in MM, 2014, pp. 457–466.
[10] Z. Wang, S. Chang, F. Dolcos, D. Beck, D. Liu, and T. S. Huang, “Brain-inspired deep networks for image aesthetics assessment,” Michigan Law Review, pp. 123–128, 2016.
[11] R. Ma, J. Liu, and C. W. Chen, “A-lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment,” in CVPR, 2017, pp. 722–731.
[12] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” SIAM Review, pp. 435–500, 2009.
[13] S. Rendle and L. Schmidt-Thieme, “Pairwise interaction tensor factorization for personalized tag recommendation,” in WSDM, 2010, pp. 81–90.
[14] E. Acar, T. G. Kolda, and D. M. Dunlavy, “All-at-once optimization for coupled matrix and tensor factorizations,” Computing Research Repository - CORR, 2011.
[15] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, “BPR: bayesian personalized ranking from implicit feedback,” in UAI, 2009, pp. 452–461.
[16] R. Datta, D. Joshi, J. Li, and J. Z. Wang, “Studying aesthetics in photographic images using a computational approach,” in ECCV, 2006, pp. 288–301.
[17] W. Luo, X. Wang, and X. Tang, “Content-based photo quality assessment,” IEEE Transactions on Multimedia, pp. 1930–1943, 2013.
[18] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka, “Assessing the aesthetic quality of photographs using generic image descriptors,” in ICCV, 2011, pp. 1784–1791.
[19] R. Salakhutdinov and A. Mnih, “Probabilistic matrix factorization,” in NIPS, 2007, pp. 1257–1264.
[20] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” Computer, pp. 30–37, 2009.
[21] Y. He, C. Wang, and C. J. Jiang, “Correlated matrix factorization for recommendation with implicit feedback,” IEEE Transactions on Knowledge & Data Engineering, pp. 1–1, 2018.
[22] C. C. Hsu, M. Y. Yeh, and S. D. Lin, “A general framework for implicit and explicit social recommendation,” IEEE Transactions on Knowledge & Data Engineering, pp. 1–1, 2018.
[23] X. He, Z. He, J. Song, Z. Liu, Y. G. Jiang, and T. S. Chua, “Nais: Neural attentive item similarity model for recommendation,” IEEE Transactions on Knowledge & Data Engineering, pp. 1–1, 2018.
[24] D. Sha, D. Wang, X. Zhou, S. Feng, Y. Zhang, and G. Yu, “An approach for clothing recommendation based on multiple image attributes,” in WAIM, 2016, pp. 272–285.
[25] V. Jagadeesh, R. Piramuthu, A. Bhardwaj, W. Di, and N. Sundaresan, “Large scale visual recommendations from street fashion images,” in SIGKDD, 2014, pp. 1925–1934.
[26] E. Simossera, S. Fidler, F. Moreno-Gomez, and R. Urtasun, “Neuroaesthetics in fashion: Modeling the perception of fashionability,” in CVPR, 2015, pp. 869–877.
[27] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver, “Mul-tiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering,” in RecSys, 2010, pp. 79–86.
[28] X. Chen, Z. Qin, Y. Zhang, and T. Xu, “Learning to rank features for recommendation over multiple categories,” in SIGIR, 2016, pp. 305–314.
[29] A. M. Buchanan and A. W. Fitzgibbon, “Damped Newton algorithms for matrix factorization with missing data,” in CVPR, 2005, pp. 316–322 vol. 2.
[30] E. Acar, T. G. Kolda, D. M. Dunlavy, and M. Morup, “Scalable tensor factorizations for incomplete data,” Chemometrics and Intelligent Laboratory Systems, pp. 41–56, 2010.
[31] P. Bhragava, T. Phan, J. Zhou, and J. Lee, “Who, what, when, and where: Multi-dimensional collaborative recommendations using tensor factorization on sparse user-generated data,” in WWW, 2015, pp. 130–140.
[32] L. Xiong, X. Chen, T. K. Huang, J. G. Schneider, and J. G. Carbonell, “Temporal collaborative filtering with bayesian probabilistic tensor factorization,” in SDM, 2010, pp. 211–222.
[33] X. He, H. Zhang, M.-Y. Kan, and T.-S. Chua, “Fast matrix factorization for online recommendation with implicit feedback,” in SIGIR, 2016, pp. 549–558.
[34] J. Ding, F. Feng, X. He, G. Yu, Y. Li, and D. Jin, “An improved sampler for bayesian personalized ranking by leveraging view data,” in WWW, 2018, pp. 13–14.
[35] W. Pan, H. Zhong, C. Xu, and Z. Ming, “Adaptive bayesian personalized ranking for heterogeneous implicit feedbacks,” Knowl.-Based Syst., pp. 173–180, 2015.
[36] W. Zhang, T. Chen, J. Wang, and Y. Yu, “Optimizing top-n collaborative filtering via dynamic negative item sampling,” in SIGIR, 2013, pp. 785–788.
[37] W. Pan and L. Chen, “Gbpr: group preference based bayesian personalized ranking for one-class collaborative filtering,” in IJCAI, 2013, pp. 2691–2697.
[38] S. Qiu, J. Cheng, T. Yuan, C. Leng, and H. Lu, “Item group based pairwise preference learning for personalized ranking,” in SIGIR, 2014, pp. 1219–1222.
[39] H. Liu, Z. Wu, and X. Zhang, “Cplr: Collaborative pairwise learning to rank for personalized recommendation,” Knowl.-Based Systems, 2018.
[40] Z. Cao, T. Qin, T. Y. Liu, M. F. Tsai, and H. Li, “Learning to rank: from pairwise approach to listwise approach,” in ICML, 2007, pp. 129–136.
[41] J. Liu, C. Wu, Y. Xiong, and W. Liu, “List-wise probabilistic matrix factorization for recommendation,” Information Sciences, pp. 434–447, 2014.
[42] X. Chen, P. Wang, Z. Qin, and Y. Zhang, “Hilbpr: A hybrid local bayesian personal ranking method,” in WWW, 2016, pp. 21–22.
[43] T. Zhao, J. Mcauley, and I. King, “Leveraging social connections to improve personalized ranking for collaborative filtering,” in CIKM, 2014, pp. 261–270.
[44] N. Murray, L. Marchesotti, and F. Perronnin, “Ava: A large-scale database for aesthetic visual analysis,” in CVPR, 2012, pp. 2408–2415.

Wenhui Yu is a Ph.D. student at Tsinghua University. His research interests span recommender system and skyline queries. He has published several papers in premier conferences including WWW and CIKM. His work on recommender system has received the Best Paper Award Honorable Mention of ACM WWW 2018.

Jian Pei is currently a Canada Research Chair (Tier 1) in Big Data Science, a Professor in the School of Computing Science at Simon Fraser University, Canada. He is one of the most cited authors in data mining, database systems, and information retrieval. Since 2000, he has published one textbook, two monographs and over 200 research papers in refereed journals and conferences, which have been cited by more than 77,000 in literature. He was the editor-in-chief of the IEEE Transactions of Knowledge and Data Engineering (TKDE) in 2013-2016, is currently a director of the Special Interest Group on Knowledge Discovery in Data (SIGKDD) of the Association for Computing Machinery (ACM). He is a Fellow of the ACM and of the IEEE.

Xu Chen is a PhD student at Tsinghua University. His research interests include machine learning, data mining, user behavior analysis. He has published over 10 papers in premier and conferences including SIGIR, WWW, CIKM, WSDM.

Li Xiong is a Professor of Computer Science and Biomedical Informatics at Emory University. She conducts research that addresses both fundamental and applied questions at the interface of data privacy and security, spatiotemporal data management, and health informatics. She has published over 100 papers in premier journals and conferences including TKDE, JAMIA, VLDB, ICDE, CCS, and WWW. She currently serves as associate editor for IEEE Transactions on Knowledge and Data Engineering (TKDE) and on numerous program committees for data management and data security conferences.

Jinfei Liu is a joint postdoctoral research fellow at Emory University and Georgia Institute of Technology. His research interests include skyline queries, data privacy and security, and machine learning. He has published over 20 papers in premier journals and conferences including VLDB, ICDE, CIKM, and IPL.

Zheng Qin is a doctoral supervisor, professor and the director of the Software Engineering and Management Research Institute and Information Institute of Tsinghua University. He conducts research that addresses both fundamental and applied questions at the formal verification of software, data mining, and computer vision. He has published over 50 papers in premier journals and conferences including WWW, SIGIR, CIKM, WSDM, PCM.

Xiangnan He is currently a professor with the University of Science and Technology of China (USTC). He received his Ph.D. in Computer Science from National University of Singapore (NUS) in 2016, and did postdoctoral research in NUS until 2018. His research interests span information retrieval, data mining, and multimedia analytics. He has over 50 publications appeared in several top conferences such as SIGIR, WWW, and MM, and journals including TKDE, TOIS, and TMM. His work on recommender systems has received the Best Paper Award Honourable Mention in WWW 2018 and ACM SIGIR 2016. Moreover, he has served as the PC member for several top conferences including SIGIR, WWW, MM, KDD etc., and the regular reviewer for journals including KDE, TOIS, TMM, TNNLS etc.