Pattern Recognition 120 (2021) 108100

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Semantic manifold modularization-based ranking for image recommendation

Meng Jian[a], Jingjing Guo[a], Chenlin Zhang[a], Ting Jia[a], Lifang Wu[a,*], Xun Yang[b], Lina Huo[c]

[a] Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[b] School of Computing, National University of Singapore, Singapore 119077, Singapore
[c] Hebei Normal University, Hebei 050024, China

ARTICLE INFO

Article history:
Received 7 February 2020
Revised 10 April 2021
Accepted 2 June 2021
Available online 12 June 2021

Keywords:
Manifold propagation
Modularization
Image recommendation
User interest

ABSTRACT

As the Internet confronts the multimedia explosion, it becomes urgent to investigate personalized recommendation for alleviating information overload and improving users' experience. Most personalized recommendation approaches pay their attention to collaborative filtering over users' interactions, which suffers greatly from the highly sparse interactions. In image recommendation, visual correlations among images that users consumed provide a piece of intrinsic evidence to reveal users' interests. It inspires us to investigate image recommendation over the dense visual graph of images instead of the sparse user interaction graph. In this paper, we propose a semantic manifold modularization-based ranking (MMR) for image recommendation. MMR leverages the dense visual manifold to propagate users' historical records and infer user-image correlations for image recommendation. Especially, it constrains interest propagation within semantic visual compact groups by manifold modularization to make a tradeoff between users' personality and graph smoothness in propagation. Experimental results demonstrate that user-consumed visual correlations play actively to capture users' interests, and the proposed MMR can infer user-image correlations via visual manifold propagation for image recommendation.

1. Introduction

The explosion of multimedia information on the Internet brings both opportunities and challenges to social multimedia computing techniques. Users' behaviors accumulate and boost the evolutions of the social multimedia environment, which makes it a survival battle for platforms to capture users' interests and provide personalized service. Towards the competition of both user service and platform development of social multimedia networks, personalized recommendation, which caters to users' interests, is becoming an actively explored field. However, it seems not easy for recommendation systems to seek certain media data to fit users' interests from such huge multimedia data. The key challenge for recommendation lies in the ”intention gap” between users' interests and overloaded data and computational burden in dealing with the real-life data. The necessity rises continuously to infer implicit interactions from users' social records for personalized recommendation. Until now, lots of excellent researchers have paid their efforts on recommendation techniques and proposed a series of elaborate recommendation models, such as collaborative filtering-based (CF) [1–4], content-based (CB) [5–7], and hybrid [8–10] recommendation models. CF models focus on unveiling users' interests from collaborative signals hidden in users' interactions, while CB ones take users' consumed multimedia contents to mine users' interests.

Besides collaborative signals and multimedia contents, we argue that in image recommendation, **visual correlations** that users consumed take strong evidence on users' interests. Since the core of recommendation systems locates in learning users' interests, we investigate the visual correlations of images immediately to infer users' interactions. Instead of the intensively studied sparse interaction graph, we notice the natural dense property in visual correlations of images and strive to propagate users' interests over the visual graph. Due to the advantage of manifold learning in information propagation, this work employs the graph smoothness of manifold learning on the visual graph to infer the correlations between users' personalized interests and visual images. As the recommendation task provides personalized services and does not require a heavy smoothness with propagation, we prefer an adaptive manner to control the propagation scale. Therefore, a modularization is equipped to the visual distribution to constrain interest propagation within decomposed visual scales for image recommendation.

This work proposes a semantic manifold modularization-based ranking (MMR) for social image recommendation. The proposed MMR captures visual correlations of global images with the topo-

[*] Corresponding author.
E-mail address: lfwu@bjut.edu.cn (L. Wu).

https://doi.org/10.1016/j.patcog.2021.108100
0031-3203/© 2021 Elsevier Ltd. All rights reserved.

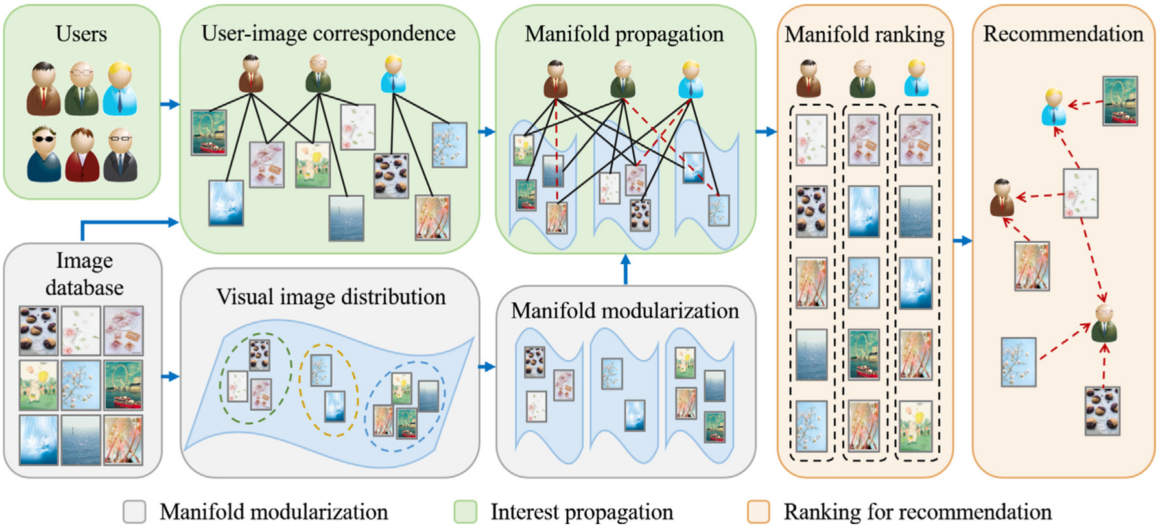


Fig. 1. Framework of the proposed manifold modularization-based ranking for image recommendation. The black solid links represent the historical user-image records. The red dashed links indicate the estimated user-image correlations by interest propagation. The red dashed arrows illustrate personalized recommendation to the target users with high user-image correlations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

logical manifold on which user-image correlations are estimated by propagating users' historical records. Fig. 1 illustrates the whole framework of the proposed MMR for image recommendation. The framework of MMR in Fig. 1 consists of three modules, including manifold modularization, interest propagation, and ranking for recommendation. Firstly, the proposed MMR captures visual correlations of images to build a visual manifold and decomposes the global manifold into several visual compact submanifolds in the modularization module. Then, the characters of users' historical records (black solid links) are propagated over the decomposed submanifolds successively to estimate users' interaction scores on unobserved images (red dashed links). Finally, unobserved images are ranked personally with the estimated interaction scores for image recommendation. The main contributions of the work are summarized as follows:

- The proposed MMR employs visual correlations of images that users consumed to reveal and infer users' interests by interest propagation over the visual graph of images instead of propagating collaborative signals over users' sparse interaction graph.
- We constrain manifold learning within visual groups adaptively to propagate users' interests and prevent bias propagated across semantics as a tradeoff between personality and propagation smoothness.
- For image recommendation, the proposed MMR introduces manifold modularization to perform interest propagation in a decomposed manner and reduce computational burden exponentially.

The experimental analysis illustrates that the proposed MMR successfully infers user-image correlations over visual distribution for image recommendation. It also verifies the active role of visual correlations in revealing users' interests. An early version of this work [16] has been presented at the Fourth IEEE International Conference on Multimedia Big Data, 2018.

2. Related works

Collaborative filtering (CF) [1,2,4,17] is a commonly used strategy in recommendation systems, which brings out recommendations by analogizing shared interactions among users. Geng et al. [17] proposed a deep feature learning model to embed heterogeneous user-image networks into a unified feature space and recommend images simply with representation interactions. He et al. [1] constructed a neural matrix factorization with an MLP model to learn user-image interactions nonlinearly. Xue et al. [2] provided an expressive item-based collaborative filtering by investigating nonlinear and high-order item relationships. Wang et al. [4] investigated high-order interactions of the user-item bipartite graph to model collaborative signals into embeddings. However, the extremely sparse user-item interactions take very limited collaborative signals. Therefore, CF-based recommendation inevitably meets incomplete interaction inference, even the cold start problem of new users or items. CF-centered systems take users' interactions as a core but ignore the **informative contents** themselves that users are interested in, which may make up the sparsity of interactions in dealing with the code start problem. To this point, Sun et al. [18] proposed a social image tag recommendation method over tag concepts by investigating tag co-occurrences. Berger et al. [19] leveraged a tag-based retrieval model to recommend images for assisting users' expression in articles. Furthermore, Zhang et al. [20] re-ranked image tags with their contents to deal with the cold start problem and proposed a user-image-tag model with a tripartite graph for personalized image recommendation. Guo et al. [9] integrated an item-item similarity matrix and user-item matrix in CF to alleviate the cold start problem in item recommendation. Current recommendation works focus on information propagation over users' interaction graph and propose various propagation strategies in the spatial or spectral domain. They still suffer significantly from the sparse interactions to support propagation.

Visual contents in user interactions are intuitively reasonable to reveal users' preferences from a view of multimedia content in almost all the online social networks. Recently, Lovato et al. proposed to learn the users' preferences using LASSO regression from image contents that users collected [5][21]. introduced visual style into clothing matching and parsing by searching visually similar styles between items. Axenopoulos et al. [22] evaluated relevances of users, tags, and conceptually similar items on which a content-based label propagation method was proposed for personalized item recommendation. Li et al. [6] proposed to leverage a variety of image features and group sparsity to learn the relationship between images and recommend image collections for users depending on their preferences. With users' interactive feedbacks, the user-image correlations are estimated for personalized ranking [23]. Rawat [24] et al. proposed a deep end-to-end model to recommend multiple tags exploiting both the content and context of the image. Wang et al. [10] constructed a hybrid recommendation model with item-based CF to generate candidate recommendation sets and content-based semantic analysis to perform recommendation. Chen et al. [8] introduced the attention mechanism of visual content into CF to construct a content-based user preference inference and recommendation. Sejal et al. [25] formed an initial set over visual features of images based on text search and employed similarity of pairwise images between query and candidates for image recommendation. Jian et al. [26] proposed a content-refined bipartite graph method that leveraged visual content relations to compensate for sparse user-image relations of the bipartite graph. You et al. [7] proposed to propagate characteristics of each tag group over image similarity of CNN features for user profiling modeling and recommendation. He et al. [27] constructed a scalable deep model to learn temporal visual evolution, which estimated the user's preferences by CNN combining with the historical records of users and the developing trend over communities. Liu et al. [28] proposed a multi-modal fused sparse topic model for user preferences and item representations with a latent factor dictionary involving visual contents for factorization-based image recommendation. Neural personalized ranking [29] investigated user-image interactions with contextual spatial, topical, and visual characters of images in representations of users and images. Social anchor units in [30] construct an anchor-unit graph of visual features and graph regularized tensor completion to estimate correlation among users, images, and tags. Wu et al. [31] designed a hierarchical attention model of entity-level and aspect-level to aggregate user latent representations with visual features to recommend images. Visual features that users consume are generally employed as auxiliary information in most prior works to augment users' or images' embeddings, while few of them investigate interactions with visual correlations.

We argue that **visual correlations** of users' historical records take intrinsic evidence to infer users' interactions. As such, we explore the role of visual correlations in users' interactions and construct a dense user-image interaction model over their intrinsic **manifold distribution** for image recommendation. Since manifold ranking [11], manifold learning-based inference [15,32–34] has achieved great success for similar tasks. Efficient Manifold Ranking [32] conducted manifold learning with an anchor graph and performed adjacency computing over the anchor graph for content-based image retrieval. K-regular nearest neighbor (k-RNN) graph [12] was built to improve the quality of graph construction in manifold-ranking. Hessian regularization is investigated in multi-view learning to model the local distribution of manifold [33]. It successfully alleviates the constant biases of Laplacian regularization and breaks its limitation of uniform distribution in discriminative learning[34]. constructed a manifold in reciprocal references and connected components to capture intrinsic dataset geometry for manifold ranking. A manifold of image superpixels was constructed to indicate the composition of users' historical records to represent users' preference for images [13,14]. Liu et al. [15] presented graph p-Laplacian to capture local distribution and effectively improved discriminative learning ability of standard manifold learning for scene recognition. The standard manifold learning scheme [11–15] is good at discriminative learning issues, which pays more attention to distinguish samples from others of different patterns. Different from distinguishing images, as a particular domain, image recommendation studies heterogeneous pairwise matching between users and images.

3. Manifold modularization-based ranking

For image recommendation, visual correlations intuitively display a piece of promising evidence to unveil users' interests and infer users' interactions. Considering the dense property of the visual graph compared to the sparse interactions graph, we investigate the role of visual correlations in modeling users' interactions with manifold propagation on the visual graph. Manifold learning [12–14] generally conducts propagation over the global manifold across semantics. When applied to a recommendation scenario, it inevitably results in interest propagation with a bias between semantic groups. It simultaneously brings a substantial computational burden on calculating with the whole manifold. Unfortunately, in the social scenario of image recommendation, users' positive interactions are extremely sparse, and the scale of images is tremendous. The incomplete and unstable interests hidden in users' historical records make the propagation across semantics not convincing, which may hurt the performance. Therefore, we introduce manifold modularization to constrain the propagation in each relatively compact semantic group while preserving personality over visual semantics. Manifold modularization-based ranking (MMR) is proposed with manifold modularization and propagation to make use of the inherent visual correlations of images for personalized image recommendation, as the framework in Fig. 1. MMR performs user-image interaction learning on modularity-based semantic sub-manifolds successively and recommends images by ranking the estimated user-image correlations, as illustrated in Fig. 1. In this section, we provide the details of manifold modularization and interest propagation of MMR for social image recommendation as follows.

3.1. Semantic manifold modularization

As mentioned above, in extremely sparse user-image interactions, manifold learning was very likely to yield misleading propagation due to complex semantic distribution. It calls for a manifold modularization strategy to constrain the propagation locally. Thanks to the great success of deep neural networks, the high-level semantic visual presentation of images, i.e., deep features by AlexNet [35], tends to distribute in relatively compact semantic groups naturally. We leverage the advantage and block interest propagation across semantic sub-manifold groups to perform manifold learning constrained within each semantic group. Learning on semantic groups individually aims to alleviate bias cross visual groups. It also favors extending learning scalability and reducing the computational burden with modularized interest propagation for image recommendation.

Inspired by [36], we assess the compactness of local visual distribution of social images by modularity to separate the natural compact groups. Depending on the change of modularity, the global manifold could be decomposed naturally into several semantic compact sub-manifolds for image recommendation. MMR treats each image as an independent sub-manifold at the beginning of manifold modularization. Given an image set $X = [x_1, x_2, \cdots, x_n]$ with a weighted visual graph $G = (X, W)$, the modularity of the global visual manifold is assessed as Equation (1), where $W = [w_{ij}]_{n\times n}$ is an affinity matrix of visual correlations w_{ij} by the inner product of the high-level semantic representations of images x_i and x_j.

$$Q = \frac{1}{2m}\sum_{i,j=1}^{n}\left[\mathrm{w}_{ij} - \frac{d_i d_j}{2m}\right]\delta\left(o_i, o_j\right)$$
$$= \frac{1}{2m}\sum_{k=1}^{K}\left[D_{in}^{S_k} - \frac{D_{ex}^{S_k\,2}}{2m}\right] \tag{1}$$

where $m = \frac{1}{2}\sum_{i,j}\mathrm{w}_{ij}$ plays to normalize modularity Q, w_{ij} indicates semantic visual correlation between images x_i and x_j, d_i=$\sum_{\mathrm{j}}\mathrm{w}_{ij}$ is the degree of image x_i, o_i indexes the assigned sub-manifold of image x_i, $\delta\left(o_i, o_j\right)$ indicates whether x_i and x_j belong to the same

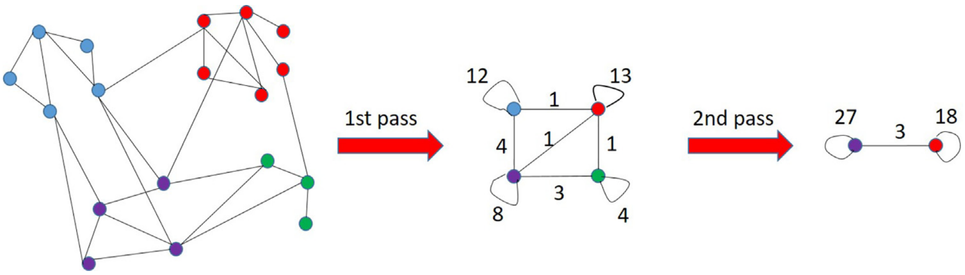


Fig. 2. The process illustration of manifold modularization, where the 1st pass measures the internal weight and the external weight of each semantic group and the 2nd pass is to fuse neighboring sub-manifolds by maximizing global modularity.

sub-manifold with $\delta\left(o_i, o_j\right) = 1$ if $o_i = o_j$ and $\delta\left(o_i, o_j\right) = 0$ otherwise, $S = [s_1, s_2, \cdots, s_K]$ is the set of sub-manifolds decomposed by modularity Q, $D_{in}^{s_k}$ is the internal weight aggregated within a specific sub-manifold s_k, $D_{ex}^{s_k}$ is the external weight aggregated from the edges incident to the sub-manifold s_k, $i, j = 1, 2, \ldots, n$, $k = 1, 2, \ldots, K$.

For each image x_i, MMR measures the change of modularity in cases of assigning x_i to its neighboring sub-manifolds. The change of modularity emerges with the gain that would occur by removing x_i from its previous sub-manifold and placing it in its neighboring sub-manifolds. The gain of modularity ΔQ when assigning an image x_i into a sub-manifold s_k is measured by Equation (2).

$$\begin{aligned}\Delta Q &= \left[\frac{D_{in}^{s_k} + d_{i,s_k}}{2m} - \left(\frac{D_{ex}^{s_k} + d_i}{2m}\right)^2\right] \\ &\quad - \left[\frac{D_{in}^{s_k}}{2m} - \left(\frac{D_{ex}^{s_k}}{2m}\right)^2 - \left(\frac{d_i}{2m}\right)^2\right] \\ &= \frac{1}{2m}\left(d_{i,s_k} - \frac{D_{ex}^{s_k} d_i}{m}\right) \end{aligned} \tag{2}$$

where d_{i,s_k} is the aggregated weight of the edges between x_i and the images in its neighboring sub-manifold s_k. Image x_i is then assigned to the sub-manifold that takes both the maximum and positive gain of modularity ΔQ. That is, if $max\Delta Q{>}0$, the image x_i is assigned to the corresponding sub-manifold. Otherwise, without positive gain, $max\Delta Q{\le}0$, x_i would be kept in its original sub-manifold. MMR repeats the above process until all the sub-manifolds that each image belongs to do not change anymore. Fig. 2 illustrates the process of manifold modularization by assigning images to their neighboring sub-manifold iteratively with global modularity. As Fig. 2, 1st pass is to measure the internal weight and the external weight of each semantic group. All the images in a sub-manifold s_k are compressed into a supernode. The weights of the edges between two images within the sub-manifold s_k are aggregated as the weight of the ring on the compressed supernode. The edge weights between sub-manifolds s_k and s_l are aggregated as the edge weight between the supernodes. In the 2nd pass, we repeat the above steps to assign supernodes into its neighboring sub-manifold until the modularity of the entire manifold does not change anymore. To the end, with a stable modularity, the global manifold has been modularized into several sub-manifolds containing local relationships of the original manifold as much as possible. With the high-level semantic representations of images, MMR assigns all the images into relatively compact semantic groups by manifold modularization, which provides a relatively pure local characteristics for the subsequent interest propagation to preserve personality and infer user-image interactions for image recommendation.

3.2. Interest propagation and ranking

As Fig. 1, manifold modularization (grey module) acts to decompose the global manifold into several independent submanifolds (blue). Then, in the interest propagation module (green), MMR propagates user-image historical records (black solid links) on each semantic compact sub-manifold s_k (blue) successively to infer global user-image correlations (red dashed links of the green module) for image recommendation. Manifold learning [12–14] is good at capturing discriminative distribution while preserving pairwise local visual correlation between images. Taking benefits of implicit local correlations in the visual manifold of images, interest propagation spreads the given sparse user-image interaction records to the corresponding submanifold s_k. It estimates dense correlations between all the images and users. This propagation aims to infer user-image correlations coincide with users' interests to images' visual characters by constraining the learning scale within visual semantic consistent groups.

Given users' historical records of images $Y = [y_{ij}]_{n\times C}$, interest propagation conducts to estimate the global user-image correlations $F = [f_{ij}]_{n\times C}$, where C is the number of users, $y_{ij} = 1$ represents the priori correlation between image x_i and user u_j, otherwise $y_{ij} = 0$, $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, C$, f_{ij} represents the estimated correlation between image x_i and user u_j through interest propagation, i.e., the user u_j's interest degree in image x_i. The proposed MMR conducts interest propagation with a standard graph smoothness term $tr(<F, LF>)$ to propagate users' interests and a constraint term $\|F - Y\|_F^2$ to preserve users' historical records Y in the estimated interaction score F. Here, L is the graph Laplacian taking manifold of visual correlations for graph smoothness-based interest propagation. Therefore, interest propagation module is formed in a tradeoff between the two terms. With manifold smoothing, MMR constructs interest propagation module as follows

$$F^* = \arg\min_F (1-\alpha)\|F - Y\|_F^2 + \alpha tr(<F, LF>) \tag{3}$$

where α plays to tradeoff between the users' specific interests of the 1st term and interest smoothness of the 2nd term, $\|*\|_F^2$ represents $Frobenious$ norm, $tr(*)$ is the trace of the matrix, and $L = I - D^{-1/2}WD^{-1/2}$ is the visual graph Laplacian matrix on the visual correlation weighted graph G constructed in Subsection 3.1, I denotes an identity matrix and D is the degree matrix of images with column sum of W. Its analytical solution can be derived as

$$F^* = ((1-\alpha)I + \alpha L)^{-1} Y \tag{4}$$

Equation (4) can be rewritten as

$$F^* = (I - \alpha S)^{-1} Y \tag{5}$$

where $S = D^{-1/2}WD^{-1/2}$. F is the inferred user-image correlations by propagating users' historical records along with the semantic compact visual groups of the manifold in Equation (3), which represents the users' underlying interest degree on images mined by interest propagation. MMR performs ultimately on F^* to sort the inferred user-image correlations in decreasing order for image recommendation. The top-N images with the highest user-image correlations are recommended to the corresponding user. We summarize the algorithm of the proposed MMR for image recommendation in Algorithm 1.

Algorithm 1 Manifold Modularization-based Ranking.

```
Input: $X = [x_1, x_2, \cdots, x_n]$ – A set of $n$ images;$Y = [y_{ij}]_{n\times C}$ – The
    given historical interactions of $C$ users with binary indicators;
Output: Personalized image recommendation lists.
1: Measure visual correlations $w_{ij}$ of images $x_i$ and $x_j$ with inner
   product over their deep visual features and construct a visual
   graph $G = (X, W)$ with the affinity matrix $W = [w_{ij}]_{n\times n}$;
2: Calculate the gain of modularity $\Delta Q$ by Equation (2) in cases of
   assigning images or image groups into their neighboring sub-manifolds;
3: Modularize the global manifold $G$ into several sub-manifolds
   $S = [s_1, s_2, \cdots, s_K]$ by maximizing the gain of modularity $\Delta Q$;
4: Propagate user-image interactions $Y$ along with each visual
   sub-manifold $s_k$ successively by manifold propagation as Equation (3)
   to estimate dense user-image correlations $F = [f_{ij}]_{n\times C}$;
5: Rank user-specific images by descending the estimated user-image
   correlations $F = [f_{ij}]_{n\times C}$;
6: return Top-N images with highest user-image correlations are
   recommended to the specific user.
```

Computational Complexity: In Algorithm 1, the major computation lays in affinity construction of **step 1**, manifold modularization of **step 2**, and manifold propagation of **step 4**. Given p images, the computational complexities of both affinity construction and modularization are $O(p^2)$, respectively, while that of manifold propagation is $O(p^3)$. Namely, in MMR without manifold modularization, i.e., $p = n$, the computational complexity of the algorithm is $O(n^3 + 2n^2)$. In the case of MMR with manifold modularization, the computational complexity is reduced a lot to $O(mp^3 + 2n^2)$, here $m < n$ is the number of sub-manifold decomposed by manifold modularization and $p << n$ is the average size of the sub-manifolds, $mp = n$. It means taking benefits of manifold modularization, the computational burden of MMR has been highly alleviated. Therefore, the proposed MMR disentangles manifold learning from its scale and successfully extends the application to personalized image recommendation in real-life social multimedia environments.

4. Experiments

In this section, we perform experiments to verify the effectiveness of the proposed MMR for image recommendation on three publicly available datasets, *Huaban* [16], *NUS-WIDE* [31], and *Pinterest* [1].

- ***Huaban*** dataset [16] is crawled from content curation social network - Huaban (http://huaban.com/), which includes images from 34 categories. The crawled dataset contains 4,737 users and 33,926 images with 1,610,984 interactions of 1.0% density, referred as *Huaban*[1].
- ***NUS-WIDE*** dataset is collected from the Flickr platform and extended by Wu et al. [31] with user-image records. NUS-WIDE contains 269,648 images from 49,545 users. As data preprocessing in [31], we filter out images and users of more than two interactions and produce a dataset of 31,460 images and 4,418 users with 761,784 interactions of 0.55% density.
- ***Pinterest*** dataset [1] from the Pinterest platform contains 12,730,502 images and 999,946 users. We filter out images of at least 30 interactions and users of more than 50 interactions, resulting in a dataset of 28,404 images and 34,235 users with 2,511,629 interactions of 0.25% density.

The characteristics of *Huaban* [16], *NUS-WIDE* [31], and *Pinterest* [1] datasets used in experiments are summarized in Table 1. The whole datasets of *Huaban* [16], *NUS-WIDE* [31], and *Pinterest* [1] illustrated in Table 1 are involved in analysing manifold propagation (Section 4.1), semantic manifold modularization (Section 4.2), recommended images vs. others in GT (Section 4.3) and experimental comparison (Section 4.4). Considering the extremely sparse interactions in *NUS-WIDE* and *Pinterest* datasets, we follow the common strategy [1,31] - for each positive image randomly sampling 100 negative images, that have no interaction with the target user, to rank the test positive images along with the negative ones. As the proposed MMR model employs manifold propagation to perform semi-supervised learning, positive images of users in the three datasets are split half as supervision and half for testing. High-level semantic representations of images are extracted by AlexNet [35] and their correlations are employed to build the visual manifold.

Widely employed protocols evaluate the performance of recommendation lists [1]: *Mean Average Precision (MAP)*, and *Normalized Discounted Cumulative Gain (NDCG)*. *MAP* takes the mean of average precisions over the top-N recommended images involving their ranked orders, which evaluate the correct recommended rate with ranking positions in lists. The higher *MAP*, the correct images take better positions at the front of the list for image recommendation. *NDCG* cumulates the position weighted relevance of recommended images in a list. Higher *NDCG* indicates that the entire ranking list takes a better image recommendation, and correct images are ranked at the front position. Due to the lack of dense user-image correspondence information, it is still hard to verify users' satisfaction on recommendation lists by *MAP* and *NDCG*. Therefore, *Content Similarity* is further introduced to analyze the performance in a visual perspective, which is conducted by evaluating the visual similarity of the high-level semantic features between the recommended image and the images in the specific users' interest records. The higher *Content Similarity*, the recommended images are more likely to attract the recommended user due to their similar semantic style and contents to the user's preference records. The recommendation lists are truncated at N for evaluation as *MAP@N*, and *NDCG@N*, $N = 1, 2, \ldots, 10$. The evaluations are provided with an average of 10 trials as follows.

The proposed MMR is further compared with several state-of-the-art models on recommendation performance including Content-Based (CB) recommendation [37], Network-Based Inference (NBI) [38], Matrix Factorization (MF) [39], Hydrid recommendation [10], Graph Convolutional Network (GCN) [40], Neural Collaborative Filtering (NeuCF) [1], Graph Convolutional Matrix Completion (GCMC) [3], Neural Graph Collaborative Filtering (NGCF) [4], Visual Neural Personalized Ranking (VNPR) [29], Hierarchical Attentive Social Contextual recommendation (HASC) [31], and Content-based Bipartite graph(CBG) [26]. All experiments are conducted on a TITAN X (Pascal) GPU server except the scalability comparison on a PC with CPU (3.6 GHz) and 32G RAM.

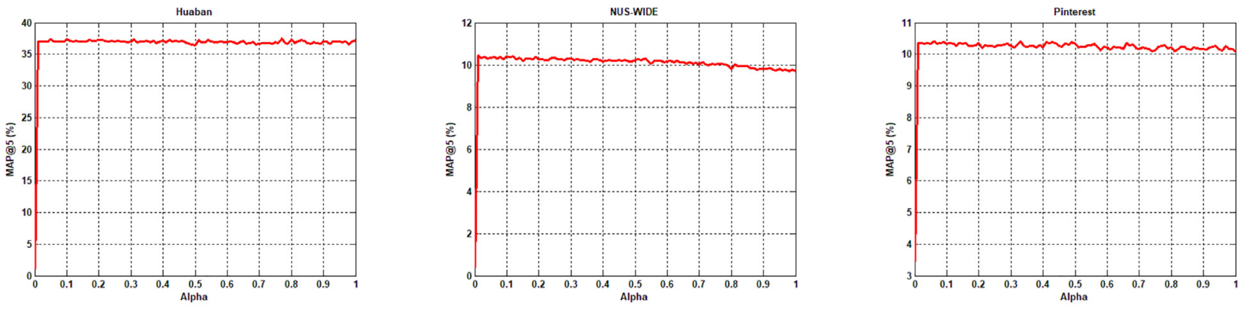
4.1. Manifold propagation in MMR

MMR builds manifold propagation as a core to spread users' interests and infer dense user-image correlations. There is a tradeoff parameter α between interest smoothness and users' specific interest. We perform experiments on the role of the parameter α in the interval of $0 < \alpha < 1$. Fig. 3 provides *MAP@5* of image rec-

[1] https://github.com/VIPL813/Huaban.com-Dataset

Table 1
Statistics of *Huaban, NUS-WIDE* and *Pinterest* datasets in experimental evaluation.

Dataset	Density	# User	# Image	# Interaction	# Median testing interaction
Huaban	1.0%	4737	33,926	1,610, 984	111
NUS-WIDE	0.55%	4418	31,460	761,784	38
Pinterest	0.25%	34,235	28,404	2,511,629	33



(a) Huaban (b) NUS-WIDE (c) Pinterest

Fig. 3. *MAP@5* of the proposed MMR for image recommendation with varying α on *Huaban, NUS-WIDE* and *Pinterest* datasets, respectively.

ommendation by MMR on *Huaban, NUS-WIDE*, and *Pinterest*, respectively, with varying α in the interval of $0 < \alpha < 1$. It shows for all the datasets the optimal performance is achieved around $\alpha = 0.01$ and the performance is decreasing with increasing of α when $\alpha > 0.01$. It means 1) more interest discriminant would derive better recommendation with smaller interest smoothness in propagation, 2) users' personalized interests should be taken more than global smoothness in image recommendation. It indicates the dominant role and necessity of personalized preference in image recommendation. This finding complies with the reality that users' personality plays a dominant role in decision-making compared to the common trend of the public. Although global smoothness helps capture users' preference among the whole space, it also tends to enhance generality by smoothing and ignoring personality. To hold the original user-image correspondence, the smoothness should not perform too much. Besides, on the previously non-existing user-image correspondence, the given user-image characteristics propagation is proportionally conducted to the whole manifold by manifold propagation, which means the ranking of these images for image recommendation would be the same with different α ideally. Without specification, we perform manifold propagation in the following experiments with $\alpha = 0.01$.

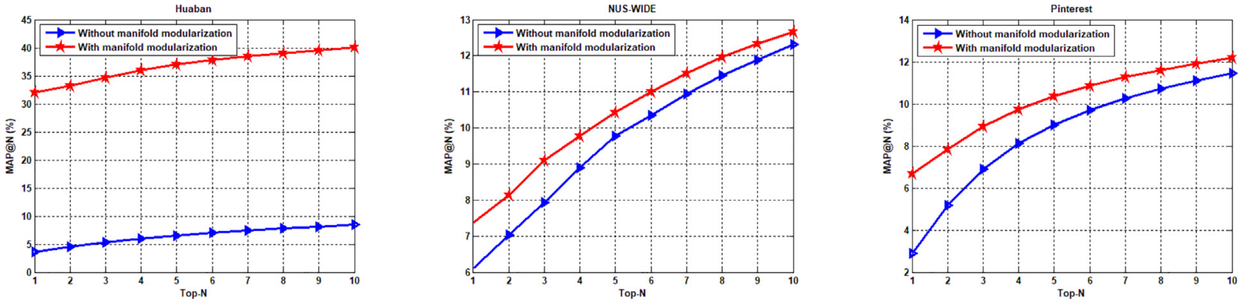
4.2. Semantic manifold modularization in MMR

To illustrate the role of manifold modularization in MMR, we conduct experiments to compare the performance of MMR with and without manifold modularization. Fig. 4 shows *MAP@N* of MMR with and without manifold modularization on *Huaban, NUS-WIDE*, and *Pinterest*, respectively, where *N* ranges from 1 to 10. The performance w.r.t. *NDCG@N* illustrates similar comparison trends. Therefore, we omit the figures of *NDCG@N*. Table 2 takes top-5 recommendation as an example and provides the performance with and without manifold modularization by *MAP@5* and *NDCG@5* on the three datasets. The performance demonstrates that manifold modularization in MMR helps improve the recommendation performance significantly. It verifies 1) the natural distribution of semantic groups on manifold takes a positive pattern for user-image correlation learning, 2) locally constrained manifold learning on modularized sub-manifolds is capable to take benefits from semantic groups of images to infer user-image correlations for image recommendation. In MMR without manifold modularization, it conducts interest propagation all over the entire visual manifold and performs interest smoothness across the global visual scale. The propagation between distinct semantic groups is too aggressive and may hurt the performance.

Besides, without manifold modularity, MMR would meet a huge computational burden to perform manifold propagation directly on the global manifold of such a real-life database like *Huaban, NUS-WIDE*, or *Pinterest*. To verify the scalability, we further perform the proposed MMR with and without manifold modularization on PC compared to those on server. Table 2 shows the performance on *Huaban, NUS-WIDE*, and *Pinterest* datasets, respectively. Due to the limited memory on PC, MMR without manifold modularization meet the out of memory issue on learning over the global manifold of about 1G data on *Huaban, NUS-WIDE*, and *Pinterest* datasets, respectively, with notation -. On the contrary, MMR with manifold modularization executes normally on the PC with comparable performance to that on the server. It implies that manifold modularization in MMR reduces the computational burden and improves the data scalability of MMR for image recommendation. Therefore, the locally constrained interest propagation with manifold modularization in MMR is of great rationality to infer user-image correlations along with visual semantics.

4.3. Recommended images vs. others in GT

Considering extreme sparsity and incompleteness of groundtruth (GT) versus dense user-image correlations, it is highly required to evaluate the probable hidden fitness of the recommendation lists for users. We also measure the semantic correlations of the recommended images and other remaining ones in GT of users' historical records to the positive images in training, comparing them with those of all images in GT and datasets by *Content Similarity*. Fig. 5 provides average *Content Similarity* on the 4,737 users of *Huaban*, 4,418 users of *NUS-WIDE*, and 34,235 users of *Pinterest*, where minimum, median, mean, and maximum of similarity are listed in order among the four groups of images. The results in Fig. 5 show that the recommended images are consistently more similar in semantic contents to users' historical interests than the others. It implies that the recommended images by MMR tend to fit users' interests more in semantic style and contents. This is because the proposed MMR employs semantic correlations in evaluating visual distribution for manifold modularization and propagating users' interests over the visual manifold. Therefore, semantic



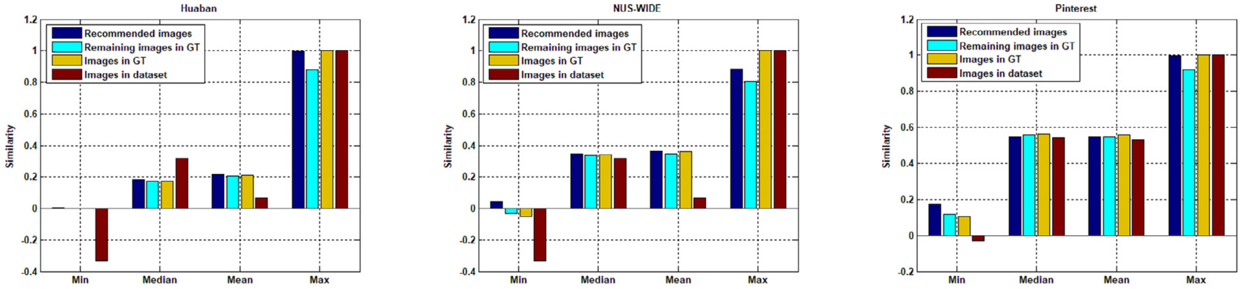
(a) Huaban (b) NUS-WIDE (c) Pinterest

Fig. 4. Performance comparison of top-N image recommendation by the proposed MMR with and without manifold modularization, where *N* ranges from 1 to 10 on *Huaban, NUS-WIDE* and *Pinterest* datasets, respectively.

Table 2
Performance comparison of the proposed MMR with and without manifold modularization by *MAP@5* and *NDCG@5* on *Huaban, NUS-WIDE* and *Pinterest* datasets, respectively.

Dataset	Metric (%)	PC		Server	
		w/o	w	w/o	w
Huaban	*MAP@5*	-	37.00	6.51	**37.01**
	NDCG@5	-	41.33	8.47	**41.41**
NUS-WIDE	*MAP@5*	-	10.40	9.77	**10.44**
	NDCG@5	-	12.77	12.69	**12.80**
Pinterest	*MAP@5*	-	10.34	9.00	**10.36**
	NDCG@5	-	12.71	7.10	**12.74**

- indicates MMR without modularization meets the out of memory issue on PC. The bold numbers mark the best results in each measure (unit: %).



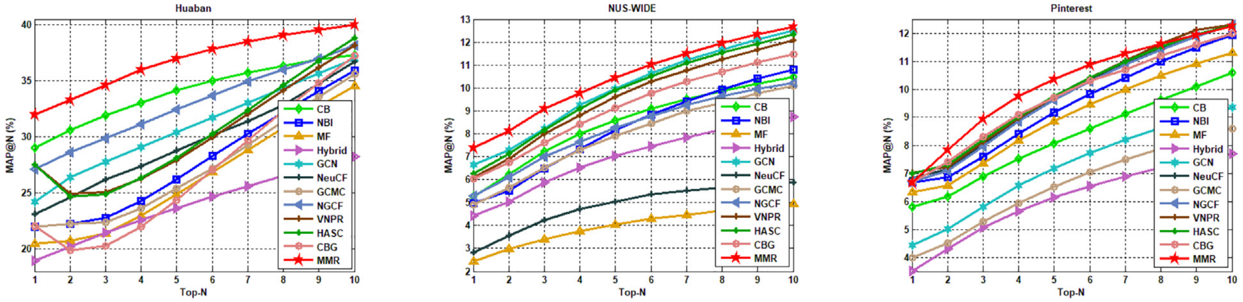
(a) Huaban (b) NUS-WIDE (c) Pinterest

Fig. 5. Average *Content Similarity* of top-10 recommended images and other remaining ones in GT of users' historical records compared with those of all images in GT and datasets.

manifold modularization and propagation in MMR help guarantee the quality of image recommendation in semantic contents.

4.4. Experimental comparison

We compare the proposed MMR with CB [37], NBI [38], MF [39], Hydrid [10], GCN [40], NeuCF [1], GCMC [3], NGCF [4], VNPR [29], HASC [31], and CBG [26] on image recommendation. The performance of the proposed MMR and all the comparisons are assessed by *MAP@N* and *NDCG@N* to quantitatively verify the effectiveness of the proposed MMR on image recommendation. Fig. 6 provides *MAP@N* of MMR compared with the comparisons on *Huaban, NUS-WIDE* and *Pinterest* datasets, respectively, where *N* ranges from 1 to 10. The performance w.r.t. *NDCG* demonstrates similar comparison trends. Therefore, we omit the figures of *NDCG@N* and show a top-5 evaluation of *MAP* and *NDCG* in Table 3. Results in Fig. 6 and Table 3 show that the proposed MMR achieves consistently better *MAP* than its comparisons do, while most interaction graph-based models like NBI [38], GCN [40], NeuCF [1], GCMC [3], and MF [39] do not perform steadily. Because these models rely greatly on the neighboring collaborative signals, which is sensitive to the interaction density of the datasets. Compared to NeuCF [1], GCMC [3], and MF [39], the performance of NGCF [4] verifies that the sensibility to interaction density could be alleviated by involving high-order collaborative signals. CB [37] performs better than GCMC [3] and MF [39], which implies the visual content is possible to uncover users' interests. To other comparisons, VNPR [29], HASC [31], and CBG [26] perform relatively better on *NUS-WIDE* and *Pinterest* datasets than that on *Huaban*. The results indicate that visual signals act positive to compensate interactions on sparse datasets like NUS-WIDE and Pinterest, while making obstacles to compromise on dense datasets like *Huaban*. Therefore, their performance on *Huaban* falls and then rises with the length of the recommendation list. With auxiliary visual signals on interactions, VNPR [29], HASC [31], and CBG [26] underperform the proposed MMR. It attributes to the adaptive propagation of users' interactions over vi-



(a) Huaban (b) NUS-WIDE (c) Pinterest

Fig. 6. Performance comparison of top-N image recommendation by *MAP@N* on *Huaban, NUS-WIDE* and *Pinterest* datasets, respectively, where *N* ranges from 1 to 10.

Table 3
Performance comparison by *MAP@5* and *NDCG@5* on *Huaban, NUS-WIDE* and *Pinterest* datasets, respectively.

Dataset	Metric (%)	CB	NBI	MF	Hybrid	GCN	NeuCF
Huaban	*MAP@5*	34.12	26.20	24.85	23.66	30.38	28.73
	NDCG@5	38.61	34.78	34.35	13.59	35.23	35.34
NUS-WIDE	*MAP@5*	8.57	8.17	4.05	7.02	9.96	5.05
	NDCG@5	10.82	10.79	5.16	8.88	12.70	6.49
Pinterest	*MAP@5*	8.07	9.17	8.86	6.15	7.16	9.70
	NDCG@5	2.45	12.01	11.44	7.80	9.41	12.56
Dataset	Metric (%)	GCMC	NGCF	VNPR	HASC	CBG	MMR
Huaban	*MAP@5*	25.43	32.41	27.91	28.13	24.33	**37.01**
	NDCG@5	33.80	37.78	36.61	37.11	33.91	**41.41**
NUS-WIDE	*MAP@5*	7.89	8.26	9.6	9.9	9.12	**10.44**
	NDCG@5	10.16	10.34	12.34	12.76	11.52	**12.80**
Pinterest	*MAP@5*	6.51	9.61	9.72	9.76	9.70	**10.36**
	NDCG@5	8.57	12.45	12.63	12.56	12.68	**12.74**

The bold numbers mark the best results in each measure (unit: %).

sual correlations in MMR. However, VNPR, HASC, and CBG depend highly on the augmented representations or graph with both visual and collaborative signals. As the definitions, *MAP* and *NDCG* evaluate the ranking list of recommended images by involving ranking order to measure the performance quantitatively. The performance means MMR recommends more positive images with better positions in ranking for image recommendation than the other models. It takes apparent advantages in recommending images along with users' specific interests. The performance demonstrates MMR implements effectively to learn users' preferences by manifold modularization and propagation. In conclusion, experimental results demonstrate that user consumed visual correlations play actively to reflect users' interests. The proposed MMR is capable to infer user-image correlations via visual manifold for image recommendation.

5. Conclusion

We have proposed a semantic manifold modularization-based ranking (MMR) for image recommendation. Instead of exploiting users' interaction graph, MMR investigates images' dense visual semantic correlations to model users' interactions for image recommendation immediately. Users' historical records are propagated along with visual correlations to learn users' personalized interests. Interest propagation is constrained within visual semantic groups by manifold modularization to preserve personality versus propagation smoothness. Experiments demonstrate the effectiveness of visual correlations in revealing users' interests and also verify semantic manifold modularization and interest propagation in estimating user-image correlations for image recommendation.

This work serves as an attempt to explore the visual correlations of images for personalized recommendation, which uncovers the rich visual semantics taking a certain role in reflecting users' interests. As correlations existing between users' interests and visual semantics, the heterogeneous inference model constructed in multi-modality is highly required. In future work, we would investigate pairwise/siamese propagations between users' interests and visual space to extend the MMR model.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Natural Science Foundation of China under Grant NO. 61702022, NO. 61802011, NO. 61976010, Beijing Municipal Education Committee Science Foundation NO. KM201910005024, Inner Mongolia Autonomous Region Science and Technology Foundation under Grant No. 2021GG0333, Beijing excellent young talent cultivation project of China under Grant NO. 20170000201-24G075, Beijing University of Technology "Ri Xin" cultivation project, NExT++ research of the National Research Foundation, Prime Minister's Office, Singapore under its IRC@SG Funding Initiative, and China Scholarship Council.

References

[1] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T.S. Chua, Neural collaborative filtering, International Conference on World Wide Web (2017) 173–182.
[2] F. Xue, X. He, X. Wang, et al., Deep item-based collaborative filtering for Top-N recommendation, ACM Trans. Inf. Syst. 37 (3) (2019).
[3] R. van den Berg, T.N. Kipf, M. Welling, Graph convolutional matrix completion, International Conference on World Wide Web (2018).

[4] X. Wang, X. He, M. Wang, Neural graph collaborative filtering, International ACM SIGIR Conference on Research and Development in Information Retrieval (2019) 165–174.

[5] P. Lovato, M. Bicego, C. Segalin, C. Perina, N. Sebe, M. Cristani, Faved! biometrics: tell me which image you like and i'll tell you who you are, IEEE Transactions on Information Forensics & Security 9 (2014) 364–374.

[6] Y. Li, T. Mei, Y. Cong, J. Luo, User-curated image collections: modeling and recommendation, IEEE International Conference on Big Data (2015) 591–600.

[7] Q. You, S. Bhatia, J. Luo, A picture tells a thousand words - about you! user interest profiling from user generated visual content, Signal Processing 124 (2015) 45–53.

[8] J. Chen, H. Zhang, X. He, et al., Attentive collaborative filtering: multimedia recommendation with item- and component-level attention, International ACM SIGIR Conference on Research and Development in Information Retrieval (2017) 335–344.

[9] C. Guo, M. Zhang, Y. Liu, S. Ma, A picture is worth a thousand words: introducing visual similarity into recommendation, International Conference on Intelligent Control and Information Processing (2017) 153–160.

[10] H. Wang, P. Zhang, T. Lu, H. Gu, N. Gu, Hybrid recommendation model based on incremental collaborative filtering and content-based algorithms, IEEE International Conference on Computer Supported Cooperative Work in Design (2017) 337–342.

[11] D. Zhou, Ranking on data manifolds, Adv Neural Inf Process Syst 16 (2003) 169–176.

[12] B. Wang, F. Pan, K. Hu, et al., Manifold-ranking based retrieval using k-regular nearest neighbor graph, Pattern Recognit 45 (4) (2012) 1569–1577.

[13] M. Jian, C. Jung, Interactive image segmentation using adaptive constraint propagation, IEEE Trans. Image Process. 25 (3) (2016) 1301–1311.

[14] L. Zhang, Y. Yao, X. Ju, et al., Massive-scale aesthetic communities learning using a noise-tolerant deep architecture, IEEE Trans Multimedia (2019) 1–11.

[15] W. Liu, X. Ma, Y. Zhou, p -Laplacian regularization for scene recognition, IEEE Trans Syst Man Cybern 49 (8) (2019) 2927–2940.

[16] T. Jia, M. Jian, L.F. Wu, Y.H. He, Modular manifold ranking for image recommendation, IEEE International Conference on Multimedia Big Data (2018).

[17] X. Geng, H. Zhang, J. Bian, T.S. Chua, Learning image and user features for recommendation in social networks, IEEE International Conference on Computer Vision (2015) 4274–4282.

[18] A. Sun, S.S. Bhowmick, J.A. Chong, Social image tag recommendation by concept matching, ACM International Conference on Multimedia (2011) 1181–1184.

[19] P. Berger, P. Hennig, D. Dummer, A. Ernst, T. Hille, Extracting image context from pinterest for image recommendation, IEEE International Conference on Smart City/socialcom/sustaincom (2016) 326–332.

[20] J. Zhang, Q. Tian, L. Zhuo, X. Liu, Y. Yang, Personalized social image recommendation method based on user-image-tag model, IEEE Trans Multimedia (2017). 1–1

[21] K. Yamaguchi, M.H. Kiapour, T.L. Berg, Paper doll parsing: retrieving similar styles to parse clothing items, IEEE International Conference on Computer Vision (2014) 3519–3526.

[22] J. Blaze, A. Asok, T.L. Roth, Content-based tag propagation and tensor factorization for personalized item recommendation based on social tagging, ACM Trans. Interact. Intell. Syst. 3 (4) (2014) 26.

[23] M. Jian, C. Jung, Y. Shen, J. Liu, Interactive image retrieval using constraints, Neurocomputing 161 (2015) 210–219.

[24] Y.S. Rawat, M.S. Kankanhalli, Contagnet: exploiting user context for image tag recommendation, ACM on Multimedia Conference (2016) 1102–1106.

[25] D. Sejal, T. Ganeshsingh, K.R. Venugopal, S.S. Iyengar, L.M. Patnaik, Image recommendation based on ANOVA cosine similarity, Int J Multimed Inf Retr 6 (5) (2017) 1–12.

[26] M. Jian, T. Jia, L. Wu, L. Zhang, D. Wang, Content-based bipartite user-image correlation for image recommendation, Neural Processing Letters (2020).

[27] R. He, J. Mcauley, Ups and downs: modeling the visual evolution of fashion trends with one-class collaborative filtering, International Conference on World Wide Web (2016) 507–517.

[28] X. Liu, M.H. Tsai, T. Huang, Analyzing user preference for social image recommendation, 2016, (????). ArXiv preprint arXiv:1604.07044.

[29] W. Niu, J. Caverlee, H. Lu, Neural personalized ranking for image recommendation, ACM International Conference on Web Search and Data Mining (ACM WSDM) (2018) 423–431.

[30] J. Tang, X. Shu, Z. Li, Y.G. Jiang, Q. Tian, Social anchor-unit graph regularized tensor completion for large-scale image retagging, IEEE Trans Pattern Anal Mach Intell (2019).

[31] L. Wu, L. Chen, R. Hong, Y. Fu, X. Xie, M. Wang, A hierarchical attention model for social contextual image recommendation, IEEE Trans Knowl Data Eng (2019).

[32] B. Xu, J. Bu, C. Chen, et al., Efficient manifold ranking for image retrieval, ACM International Conference on Research and Development in Information Retrieval (SIGIR) (2011) 525–534.

[33] W. Liu, D. Tao, Multiview hessian regularization for image annotation, IEEE Trans. Image Process. 22 (7) (2013) 2676–2687.

[34] D.C. Pedronette, F.M. Goncalves, I.R. Guilherme, et al., Unsupervised manifold learning through reciprocal Knn graph and connected components for image retrieval tasks, Pattern Recognit (2018) 161–174.

[35] A. Krizhevsk, I. Sutskever, G. Hinton, Imagenet classification with deep convolutional neural networks, Adv Neural Inf Process Syst (2012).

[36] V.D. Blondel, J.L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of community hierarchies in large networks, J. Stat. Mech: Theory Exp. 10 (2008) 10008.

[37] M.J. Pazzani, D. Billsus, Content-based recommendation systems, Adaptive Web 4321 (2007) 325–341.

[38] T. Zhou, J. Ren, M. Medo, Y.C. Zhang, Bipartite network projection and personal recommendation, Physical Review E Statistical Nonlinear & Soft Matter Physics 76 (2007) 046115.

[39] S. Rendle, C. Freudenthaler, Z. Gantner, BPR: Bayesian personalized ranking from implicit feedback, Uncertainty in Artificial Intelligence (2009) 452–461.

[40] M. Niepert, M.H. Ahmed, K. Kutzkov, Learning convolutional neural networks for graphs, International Conference on Machine Learning (2017) 2014–2023.

Meng Jian received the B.S. and Ph.D. degrees from Xidian University, China, in 2010 and 2015, respectively. She is currently an associate professor with the Faculty of Information Technology, Beijing University of Technology, China. She is also a Research Scholar with School of Computing, National University of Singapore, Singapore, from Nov. 2018 - Nov. 2019. She has been awarded Beijing Excellent Young Talent in 2017 and "Ri xin" Talents of Beijing University of Technology in 2018. Her main research interests include pattern recognition, image understanding and social media computing.

Jingjing Guo received the B.E. degree in Electronic Information Engineering from Beijing University of Technology, China, in 2020. She is currently pursuing her M.E. degree in Beijing University of Technology, China. Her research interests include heterogeneous network representation and recommendation systems.

Chenlin Zhang received the B.S. degree in Electronic information science and technology from Lanzhou University of Technology, China, in 2019. He is currently pursuing the M.S. degree in Beijing University of Technology, Beijing, China. His research interests include pattern recognition and recommendation system.

Ting Jia received her M.E. degree from Beijing University of Technology, Beijing China, in 2019. She received her B.S. degree in communication engineering from Tangshan University, China, in 2012. Her research interests include pattern recognition and recommendation system.

Lifang Wu received her B.E. and M.E. degree from Beijing University of Technology, Beijing, China, in 1991 and 1994, respectively. She received her Ph.D. degree of pattern recognition and intelligent system from BJUT in 2003. She is now a professor with the Faculty of Information Technology, Beijing University of Technology. She has published over 50 referred technical papers in international journals and conferences of image/video processing, pattern recognition. Her research interests include image/video analysis and understanding, face detection and recognition, face encryption. She is a senior member of Chinese Institute of Electronics.

Xun Yang received the Ph.D. degree in 2018 from School of Computer and Information at Hefei University of Technology, China. He is currently a Research Fellow in NExT++ center, School of Computing at National University of Singapore. His research interests include information retrieval, computer vision, and multimedia information processing.

Lina Huo received the B.S degree in electronic information engineering from Shijiazhuang Railway Institute, Shijiazhuang, China, in 2006. In 2016, she received the Ph.D. degree from Xidian University, Xi'an, China. She joined the faculty of College of Mathematics and Information Science, Hebei Normal University in 2016. Her research interests include machine learning, image saliency, object detection, and image representation.