One of a Kind: User Profiling by Social Curation

Xue Geng†, Hanwang Zhang†, Zheng Song§, Yang Yang‡, Huanbo Luan† and Tat-Seng Chua†
†School of Computing, National University of Singapore
‡University of Electronic Science and Technology of China
§Visenze Pte. Ltd.
{snownus, hanwangzhang, sirius.szh, dlyyang}@gmail.com
{luanhb, chuats}@comp.nus.edu.sg

ABSTRACT

Social Curation Service (SCS) is a new type of emerging social media platform, where users can select, organize and keep track of multimedia contents they like. In this paper, we take advantage of this great opportunity and target at the very starting point in social media: user profiling, which supports fundamental applications such as personalized search and recommendation. As compared to other profiling methods in conventional Social Network Services (SNS), our work benefits from the two distinguishable characteristics of SCS: a) organized multimedia user-generated contents, and b) content-centric social network. Based on these two characteristics, we are able to deploy the state-of-the-art multimedia analysis techniques to establish content-based user profiles by extracting user preferences and their social relations. First, we automatically construct a content-based user preference ontology and learn the ontological models to generate comprehensive user profiles. In particular, we propose a new deep learning strategy called multi-task convolutional neural network (mtCNN) to learn profile models and profile-related visual features simultaneously. Second, we propose to model the multi-level social relations offered by SCS to refine the user profiles in a low-rank recovery framework. To the best of our knowledge, our work is the first that explores how social curation can help in content-based social media technologies, taking user profiling as an example. Extensive experiments on 1,293 users and 1.5 million images collected from Pinterest in fashion domain demonstrate that recommendation methods based on the proposed user profiles are considerably more effective than other state-of-the-art recommendation strategies.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Sociology*

Keywords

Social Curation; User Profiling; Recommendation; Pinterest

MM'14, November 3–7, 2014, Orlando, Florida, USA.
Copyright 2014 ACM 978-1-4503-3063-3/14/11 ...$15.00.
http://dx.doi.org/10.1145/2647868.2654950 .

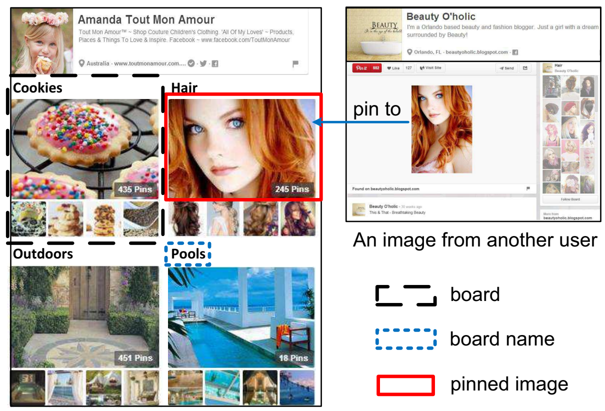


Figure 1: An example of "pin" in Pinterest. A user can curate image bundles called "boards". The curation is done by "pinning" images from other users. Here, the user pins a hairstyle image from another user into her own board named "hair".

1. INTRODUCTION

Social Network Service (SNS), through which people can create, disseminate, and consume information, has evolved themselves while revolutionized our lives over the past 15 years[1]. To date, a large variety of SNSs have thrived on the Internet, focusing on retail market (*e.g.*, Amazon), friendship (*e.g.*, Facebook), movie review (*e.g.*,IMDb), photo sharing (*e.g.*, Flickr), and so on. It is widely acknowledged that social media offers us valuable opportunities in both academia and industry [6].

In this paper, we are interested in user profiling, which is the starting yet fundamental point of SNS [3] in helping to customize large scale web contents to specific users. Here, profiling is the process of establishing users' profiles of interest from the spontaneous data associated with users on SNS, with the aim of capturing the online behavior of the users, including their interest and preferences. Without doubt, user profiling is an indispensable direction in social media, especially for the recommendation-based applications such as viral marketing and e-commerce [14]. User profile can be extracted from explicit or implicit sources. The *explicit* user profile is provided by users during the registration to some

[1]The first online social network, FriendsReunited was launched in 1999, founded in Great Britain to reunite past school pals.

services, and it is often incomplete and inaccurate. *Implicit* profiling is generally content-based, which has been shown to be a useful enhancement based on user relations (*e.g.*, followers or friends) and user-generated contents (*e.g.*, reviews and uploaded/shared photos, videos), which are often multimedia in nature [1]. In fact, the major interest of SNS is rapidly shifting from text-based contents to multimedia[2].

However, content-based user profiling by analyzing the multimedia data in conventional SNSs (*e.g.*, Facebook, Flickr) is a challenging task since the philosophy behind them is rooted in the social connections between the contents creators, *i.e.*, the users, but not the contents themselves. On one hand, the user-user and user-content connections in these conventional networks are *abundant but less informative* to derive personalized user profiles. For example, the friendship of two users does not necessarily suggest that they share the same preference, and the intricate "like" or "dislike" user-content links over the network are extremely diverse in topics, giving rise to difficulties in analyzing the exact user preference. On the other hand, the diverse and noisy user-generated contents prevent us from learning content-based models of user preference, leaving them as the "dark matters" of the "social universe" [5]. Therefore, conventional SNSs are unable to handle the increasingly multimedia nature of users' interactions and experiences.

Recently, Social Curation Service (SCS) has emerged as a new type of social network scheme, attracting many social network users [12]. As compared to the conventional SNSs, SCSs involve human process of remixing social media for the purpose of further consumption. In particular, Pinterest[3], which is the most popular SCS registered by over 70 million users, is a new emerging photo sharing social networking site that allows users to select, organize and keep track of images they like. Pinterest is a "pinboard-style" image sharing social network. Its main innovation is to encourage users to collect and share interesting things in a categorized way. As illustrated in Figure 1, Pinterest innovates a notion called "*Pin to Board*", where users can '*pin*' or '*repin*' items they like into their own "*boards*". The key operation "pin" is to select a photo or video from external websites or another users' pin boards. The boards are bundles of pinned multimedia contents of various interest such as "Animals", "Arts", "Education" and "Fashion". For example, a user can have many bundles named "Cookies", "Outdoors", and "Pools" shown in Figure 1. In this way, social connections are encoded by pins, *e.g.*, users cannot directly send private or public messages to each other and the only social activity is to *like* a pin, *comment* on a pin or *repin* someone's pin into her own boards. Today, many conventional SNSs are inspired by this interesting feature of social curation, such as Flickr's "*add-to-gallery*".

Motivated by the promising outlook of social curation, we attempt to establish high-quality user profiles based on such new social media platforms, *i.e.*, SCSs, with the aim of advancing fundamental social applications such as recommendation. Specifically, our user profiling approach is superior to traditional ones for two key reasons:

Organized vs. Unorganized Contents. Unorganized multimedia contents in conventional SNSs are visually and semantically noisy and diverse, and are thus hard to analyze and exploit even with the state-of-the-art multimedia annotation techniques [11]. In contrast, SCSs contain a considerable amount of manually organized and maintained contents. For example, images in a curated bundle (*e.g.*, the board in Pinterest or the gallery in Flickr) are very focused on the same semantics as shown in Figure 1. Such organized multimedia contents offer us high-quality human labeled training data for multimedia modeling. Moreover, we are able to mine a large amount of curated bundles of user interest to build an content-based ontology to further structuralize the data, resulting in more personalized and accurate user profiles.

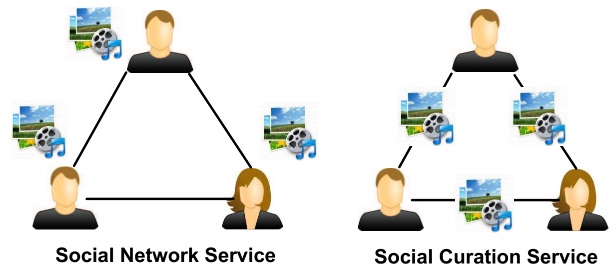


Figure 2: The illustration of the user-centric SNS and the content-centric SCS. Although SNS contains user-generated content, the philosophy behind it is limited to user-user interactions. Alternatively, SCS encourages user interactions at the content level.

Content-centric vs. User-centric Network. As aforementioned, conventional user-centric SNSs are not optimized to create comprehensive user profiles based on user-generated contents. Alternatively, as illustrated in Figure 2, content-centric SCSs are advantageous in reliable social cues on user preference. In particular, user curation through multimedia contents helps to encode multi-level content-content connections, which are expected to pinpoint the user preference in terms of the contents generated by the user. Such connections between two images include: a) *user-level* connection, where the strength indicates how many users have pinned the two images; b) *bundle-level* connection, where the strength indicates how many bundles share the two images; and c) *content-level* connection, where the strength suggests the similarities between the two images. In particular, the first two connections are expected to unravel the diverse user interest hidden in the contents. For example, if two images are only shared by few users (or bundles), the connection between them rarely suggests similar user interest. However, if they are shared by many users (or bundles), they tend to be very likely referring to the same interest. Our user profiling method can leverage rich information to refine the imperfect content-based profile models.

The overview of the proposed user profiling approach based on an example of SCSs, Pinterest, is illustrated in Figure 3. First, we collect multimedia contents curated by users, *i.e.*, the images in bundles as well as the associated user interest description like bundle names and tags. Due to user curation, the collected data are of high-quality and focused according to the user interest. Second, we propose an automatic ontology construction method to structuralize the curated images onto an ontology. The construction is done by pruning an expert ontology, *i.e.*, Wikipedia Category,

[2]http://www.kpcb.com/insights/2013-internet-trends
[3]http://www.pinterest.com

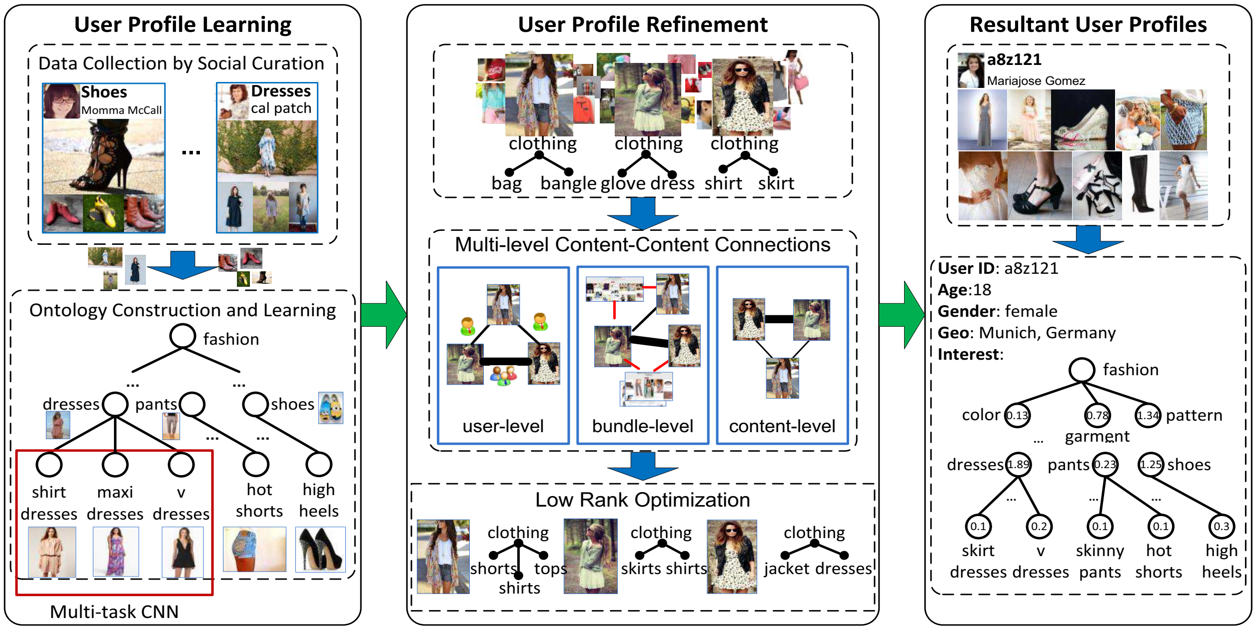


Figure 3: The overview of the proposed user profiling by social curation.

to the desired user interest in a specific domain, such as fashion. Third, based on the constructed ontology, we are able to learn content-based models to generate ontological user profiles, which are more comprehensive and personalized than the traditional text-based profiles. In particular, we propose a novel multi-task convolutional neural network (mtCNN) in order to leverage both the relatedness of sibling items of user interest and the cutting-edge advances in high-performance visual modeling. Fourth, we further propose a low-rank recovery framework to further refine the generated user profiles by the ontological profile models, exploiting the rich user-level, bundle-level and content-level social relations offered by social curations. Therefore, the resultant user profiles are expected to retain: a) the interest of user; b) the interest of user-curated bundles; and c) the semantic affinities with respect to the ontology, supporting effective fundamental social media applications such as recommendation. Experimental results on 1,239 users and 1.5 million images collected from Pinterest in fashion domain demonstrate the effectiveness of the proposed user profiling method as compared to other state-of-the-art methods in terms of recommendation.

With the rapid growing popularity of social curation services, we believe that our research is a pioneering work on content-based social curation analysis, with the following contributions:

- We propose a novel content-based user profiling method based on social curation. Our work concentrates on exploring how social curation can help in content-based social multimedia analysis

- We present a user profiling framework on how to exploit the rich social information in SCS. This framework is fully automated and can be extended to general visual-oriented domain. In this paper, we use the fashion domain in Pinterest as an example.

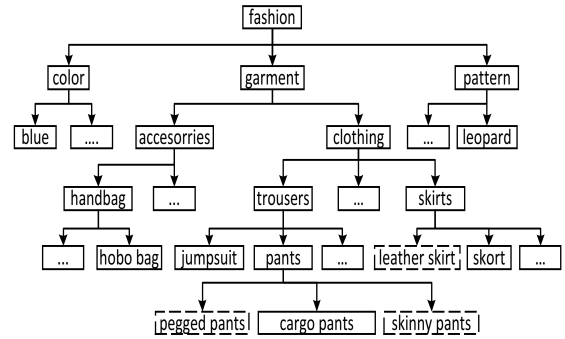
- Through extensive experiments on fashion recommendation in social media, we demonstrate that our user profiling approach considerably outperforms other state-of-the-art methods.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 describes the user profile learning process. Section 4 illustrates the process of user profile refinement. Experimental results and analysis are reported in Section 5, followed by conclusions in Section 6.

2. RELATED WORK

User profiling aims to establish user profiles by obtaining, extracting and representing the characteristics and preferences of users [34]. Current studies on user profiling often analyze either user behaviors (*e.g.*, click streams [32] and posts [28]) or user generated data to model users. From the content-based user profiling point of view, traditional user profiling methods applied either hand-crafted meta-descriptors or a set of latent features by factorizing a user's registered profile data, without considering homophily [29]. Even some researchers considered exploiting collaborative filtering techniques to profile user interest, the social connections are often sparse and noisy, leading to imperfect user profiles. On the other hand, some researchers attempted to exploit user generated contents to profile users. For example, Abel *et al.* [1] enriched user-generated text, *i.e.*, twitter tweets to construct semantic user profiles. Cai *et al.* [4] proposed to build a tag-based user profile for personalized search. Some researchers also attempted to exploit cross-SNS contents to extract comprehensive user profiles [2]. However, traditional content-based user profiles rarely build upon multimedia contents. Even if they do, they are generally unsatisfactory due to the sparse and noisy user-generated data. Alternatively, we propose to use the emerging social network platforms, the social curation services (SCS), to generate content-based user profiles.

SCSs encourage users to curate bundles of multimedia contents from diverse sources, and then re-organize them to show own interest[4] [20, 21]. As introduced by Kimura [12], the key advantage of social curation is the valuable human effort involved in organizing web contents, leading to well-organized contents and content-centric network, which is different from traditional user-centric network. For example, some researchers have worked on exploiting the Flickr groups for image annotation [25, 18] , but the work only focused on the characteristics of the well organized contents without content-centric network; while Kimura's work is just a preliminary study on the potential of social curation in image annotation. Besides, Katsuhiko *et al.* [11] showed that social curation is a promising data source for automatic image understanding and mining in SNSs compared to other state-of-art computer vision methods. Morever, Katsuhiko *et al.* used only the textual meta-data of images, but not utilizing the full power of multimedia and content-centric network. Research on social curation is still in its infancy and thus requires a lot of explorations, especially on some fundamental issues. In this work, we proposed to deploy the power of social curation and state-of-the-art visual model to establish comprehensive content-based user profiles, to support social media applications such as recommendation [22, 15, 7].



(a) Profile Ontology

#	Statistics
leaf nodes	427
total nodes	464
average samples of leaf nodes	546
least/largest leaf samples	177/11,442
least/largest depth	3/6

(b) Ontology Statistics

Figure 4: (a) The automatic constructed profile ontology in the fashion domain. The dashed boxes denote the automatically augmented concepts. (b) The ontology statistics based on user-generated contents collected from Pinterest.

3. USER PROFILES LEARNING

Social curation service (SCS), by nature, contains high-quality images in bundles curated by users. Here, "high-quality" means the semantics of images are highly constrained and relevant by the user-provided tags. This gives us a great opportunity to develop well-generalized content-based models to predict the semantics of images, in our cases, the items of user interest.

3.1 Profile Ontology Construction

We use an ontology to organize the items of user interest and their relationships in fashion domain from general to specific, as it has been widely shown to be effective in integrating human knowledge of the domain and data distributions to improve the modeling of visual semantics [31].

After harvesting the user-curated interest concepts for pinned images such as comments and bundle names, we want to automatically generate a profile ontology $\mathcal{O} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v\}$ is a set of concepts such as "leopard", "dresses" and an edge in $\mathcal{E}$ is an ordered pair of concepts in $\mathcal{V} \times \mathcal{V}$. Now, we introduce how to automatically construct the ontology $\mathcal{O}$ by mining the contents in Pinterest. First, we exploit Wikipedia Categories to build a preliminary ontology based on pre-defined general user interest. This ontology is rooted from three general concepts: "color", "garment" and "pattern", voted by the most general WordNet concepts which are hypernyms of the user provided fashion words. However, it is hard to adapt to the real user interest distribution of the user-curated data, since a) some concepts are outdated such as "polonaise", which are missing from the user-curated data; and b) high-frequency concepts such as "V-dresses", "sleeveless dresses", on the other hand, are missing from this ontology. Therefore, we should prune the Wikipedia ontology to the user interest on demand.

Specifically, to remove the outdated concept, we consider the concepts with low term frequency (*e.g.*, less than 100 times) derived from around 800,000 user-generated items as the outdated ones. Also, we need to add high-frequency concepts into this ontology. This is not a trivial task since it is challenging to find which concept in the ontology is most semantically related to a given high-frequency concept. Here, we propose a novel method for augmenting the ontology with out-of-vocabulary concepts. Suppose we want to add a high-frequency concept h onto the existing ontology $\mathcal{O}$, the key is to find the most probable semantic path from top to bottom and then add h as a sibling of a concept v if v is most semantically similar to h among others along the path. In order to numerically calculate the most probable semantic path, we need to transform concept words into numeric vectors. Here, we use Word2Vec [19] to transform a concept into a 300-D vector, retaining its semantic meanings. Then, we use all the concepts in $\mathcal{V}$ to sparsely represent h as

$$\arg\min_{\mathbf{a}} \|\mathbf{h} - \mathbf{V}\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_1, \quad (1)$$

where $\mathbf{h}$ is the 300-D vector of item h, $\mathbf{V}$ is a dictionary matrix which is arranged by putting vectors of $\mathcal{V}$ column-wisely, $\mathbf{a}$ is a sparse coefficient vector and λ is a trade-off parameter. By doing this, each concept of the ontology $\mathcal{O}$ is assigned a value according to the sparse codes of h [30]. Therefore, we can find the most probable semantic path with the largest sum of sparse codes to achieve the target concept.

As a result, we obtain a comprehensive user profile ontology as shown in Figure 4.

[4]http://scobleizer.com/2010/03/27/the-seven-needs-of-real-time-curators/

3.2 Profile Ontology Learning

Content-centric SCSs offer well-organized contents which closely relate to user interest. In other words, we are able to collect high-quality training images for every concept in the constructed profile ontology. For learning the model of each concept in the profile ontology, we want to map the images curated by users onto this profile. For example, given an image of "cargo pants", we expect to *visually* reason like: $garment \rightarrow trousers \rightarrow pants \rightarrow cargopants$. As compared to the flat "bag-of-bundles" image organizations in Pinterest, this hierarchical reasoning gives richer semantic interpretations of user interest. In order to achieve this, for each concept v in the profile ontology $\mathcal{O}$, we need to learn a classification model that predicts whether an image belongs to v. Let us start with looking for the training samples of v. Trivially, the images of the concept itself will be the positive samples. Moreover, we consider positive samples of v's siblings as negative samples of v. This myopia way of training is shown to be effective in hierarchical visual task [17]. However, this training strategy suffers from the "error propagation" problem, *i.e.*, the models of v and its siblings are incapable of rejecting the classification errors propagated from higher-level unseen concepts.

In order to alleviate such propagated errors, we expect the model of every concept in the ontology to be as accurate in prediction as possible. To achieve this, we propose to adopt the Multi-task Learning (MTL) framework [9] for jointly learning the models of a concept and its siblings. It has been shown that MTL improves the prediction performance on multiple, different but related, learning problems through shared parameters or representations. In our case, the tasks of learning a concept and its siblings are related since sibling samples are visually similar. For example, "V-dresses" and "strapless dresses" under "dresses" share similar visual cues. Formally, without loss of generality, we only consider a set of sibling concepts $\{v_1, ..., v_M\}$, which share the same parent. Given training images $\{(\mathbf{x}_i, y_i)\}$, where $\mathbf{x}_i$ and y_i are the feature and label of the i-th image in any v_m, respectively. The objective of the MTL is

$$\min_{\mathbf{w}_0,...,\mathbf{w}_M} F\left(\mathbf{w}_0, ..., \mathbf{w}_M\right) = \\ -\sum_{m=1}^{M} \sum_{i \in \mathcal{I}_m} \log P\left(y_i = m | \mathbf{x}_i; \mathbf{w}_0, \mathbf{w}_m\right) + \lambda \sum_{m=0}^{M} R(\mathbf{w}_m) \tag{2}$$

where $\mathbf{w}_0, ..., \mathbf{w}_M$ are the trainable parameters for M tasks, $\mathcal{I}_m$ is the set of training image indices of v_m, $P(y_i = m|\mathbf{x}_i)$ is a softmax function against other labels $y_i \neq m$ and λ is the trade-off parameter of the regularizer $R(\cdot)$, *e.g.*, ℓ_2-norm regularizer. Particularly, $\mathbf{w}_0$ is the shared parameters of the M tasks, namely, in the last two-layer connection, $\mathbf{w}_0$ is added to $\mathbf{w}_m$ for each task m.

Recent advances in computer vision have shown that deep Convolutional Neural Network (CNN) can learn useful features that outperform the hand-crafted ones [13]. Therefore, we propose a novel multi-task CNN (mtCNN) deep architecture that jointly learns the features and parameters for the tasks. As illustrated in Figure 5, we have M CNNs for each task to learn the feature $\mathbf{x}_i \leftarrow \phi(\mathbf{x}_i; \widetilde{\mathbf{w}}_m)$, where $\phi(\mathbf{x}_i; \widetilde{\mathbf{w}}_m)$ is the output of the m-th CNN (cf. Section 5.1.2 for the details of the mtCNN architecture) and $\widetilde{\mathbf{w}}_m$ is the trainable parameters. Denoting the the overall parameters of mtCNN as $\mathcal{W} = \{\mathbf{w}_0, \mathbf{w}_m, \widetilde{\mathbf{w}}_m\}_{m=1}^{M}$, then the stochastic gradient

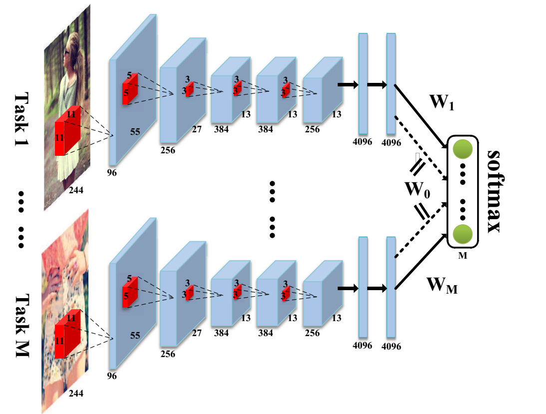


Figure 5: The illustration of the proposed mtCNN. There are M independent CNN pathways for the M tasks. These CNNs share a common parameter $\mathbf{w}_0$ when they eventually feed-forward into the softmax classification layer.

descent update rule of $\mathcal{W}$ in the k-th iteration for solving mtCNN is

$$\begin{cases} \Delta_{k+1} = 0.9 \cdot \Delta_k - 1.5e^{-4} \cdot \eta \cdot \mathcal{W}_k - \eta \cdot \frac{\partial F}{\partial \mathcal{W}_k}, \\ \mathcal{W}_{k+1} = \Delta_{k+1} + \mathcal{W}_k, \end{cases} \tag{3}$$

where Δ is the momentum variable [23], η is the learning rate which is adaptive to the objective function value.

Now we can represent any image in the user-curated bundles as $\mathbf{p}$, where the i-th value p_i is the model output of concept v_i in ontology $\mathcal{O}$. By averaging the $\mathbf{p}$ of all the images, we can eventually obtain the user profiles in a personal hierarchical representation in which the value of each concept node shows the user's interest.

4. REFINEMENT OF USER PROFILES BY SOCIAL CURATION

So far, the above user profiles are only based on visual models, which are insufficient to accurately predict user interest in terms of concepts in the ontology. In this section, we propose to refine the image representations via the multi-level social cues. Since our user profile comes from the integration of all the image representations; through refining the image representation, we can thus refine the user profiles.

4.1 Formulation

We use graph links to model the various types of relations evident from rich social information of curated images. We observe that there are three levels of key relations in the content-centric social curation network. As we will detail soon, these levels of connections play an important role in regularizing the user interest depicted in images.

User-level. This level's connection origins from the fact that "Great minds think alike". For example, if users A, B and C share images i and j simulataneously, then images i and j might be similar. Therefore, images are connected if two or more users have curated them. Formally, we have

$$S_{ij}^{u} = \begin{cases} n, \text{ if } n \text{ users share images } i \text{ and } j, \\ 0, \text{ no users share them.} \end{cases} \tag{4}$$

The strength of user-level link S_{ij}^u indicates how many users consider images i and j belong to the same interest.

Bundle-level. This level's connection is similar to the user-level connection since each bundle often represents one kind of a user's interest. Therefore, at this level, images are connected if two or more bundles include them,

$$S_{ij}^b = \begin{cases} n, & \text{if } n \text{ bundles include images } i \text{ and } j, \\ 0, & \text{no bundles share them.} \end{cases} \quad (5)$$

The strength of bundle-level link S_{ij}^b suggests how many bundles would be curated by users to pin images i and j to the same interest.

Content-level. This level includes two types of image content links: visual link and semantic link. The visual link is based on the visual similarities while the semantic link is based on the hierarchical semantic similarities between two images. Formally, we have

$$S_{ij}^v = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\rho^2}\right), \quad S_{ij}^h = \mathbf{p}_i^T \mathbf{H} \mathbf{p}_j = \sum_{k,l} \mathbf{p}_i^k H_{kl} \mathbf{p}_j^l, \quad (6)$$

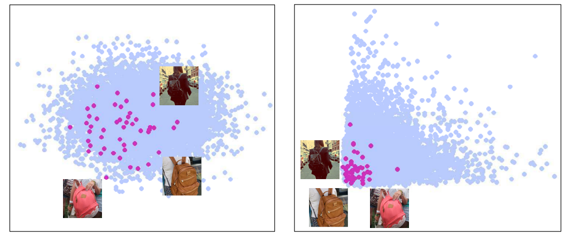
where ρ is a predefined radius set to the standard variance of the feature norms, $\mathbf{x}$ and $\mathbf{p}$ are the visual and hierarchical semantic representations of images, respectively. The matrix $\mathbf{H}$ can be derived by measuring the closeness of concept relations in the ontology. For instance, let $H_{kl} = \xi(\pi(k,l))$, where $\pi(k,l)$ is the lowest common ancestor of concepts k and l, and $\xi(\cdot)$ is some real function that is non-decreasing going down the hierarchy, *i.e.*, the lower the shared ancestor, the more similar concepts k and l are.

Now, we can combine the above links in order to refine the existing hierarchical representations of images, which are the basis to establish user profiles. Denote the hierarchical representations of all the training images before refinement as $\mathbf{P}$ and those after refinement as $\mathbf{R}$, which is the goal we are pursuing. In particular, we assume that the original $\mathbf{P}$ is noisy and the desired $\mathbf{R}$ is a low-rank recovered noise-free matrix. Intuitively, each column vector of the low-rank matrix $\mathbf{R}$ denotes the hierarchical item representation of an image. Intuitively, due to the relations between concepts in the ontology, the concept "cargo pants" should imply that concepts "pants" and "trousers" are along the semantic path. This indicates, from the viewpoint of linear algebra, that "cargo pants" could be located in a subspace spanned by those concepts along the path, imposing a low-rank nature of the matrix $\mathbf{P}$.

Therefore, by jointly considering the aforementioned multi-level social relations and the low-rank prior, the formulation of the proposed profile refinement objective is

$$\begin{aligned} \min_{\mathbf{R}} J(\mathbf{R}) &= \|\mathbf{P} - \mathbf{R}\|_F^2 + \alpha\|\mathbf{R}\|_* \\ &\quad + \beta \mathrm{trace}\left\{\mathbf{R}^T\left(\mathbf{L}^u + \mathbf{L}^b + \mathbf{L}^v + \mathbf{L}^h\right)\mathbf{R}\right\}, \end{aligned} \quad (7)$$

where α and β are trade-off parameters, $\mathbf{L}^u$, $\mathbf{L}^b$, $\mathbf{L}^v$, and $\mathbf{L}^h$ are the graph Laplacians of the corresponding graphs in Eq. (4) to (6). For example, $\mathbf{L}^u = \mathbf{D}^u - \mathbf{S}^u$, where $\mathbf{D}^u$ is a diagonal matrix with the i-th entry as $\sum_j S_{ij}^u$. Such graph regularized terms impose the low-rank pursuit of $\mathbf{R}$ to be consistent with the multi-level social connections. The nuclear norm $\|\mathbf{R}\|_*$ is a convex surrogate for matrix rank [27], whose convexity allows an effective optimization for its solution. Next, we show how to solve the formulation in Eq (7).



(a) Original User Profiles (b) Refined User Profiles

Figure 6: Illustration of the effectiveness of the proposed profile refinement method. The gray blue points are images represented by user profiles, and the red ones correspond to "bag" images. All these points are visualized by using PCA mapped into a 2-D space. (a) Before refinement, images of the same interest are scattered. (b) After refinement, images of the same interest are clustered.

4.2 Solution

When we investigate the formulation in Eq (7), we find that the profile refinement problem is a convex optimization problem. Therefore, there is a guarantee to obtain a global minimum. However, it does not have a closed-form solution. Fortunately, this problem can be solved by the Proximal Gradient method [24], which uses a sequence of quadratic approximations of the objective function $J(\mathbf{R})$ in order to derive the optimal solution.

We define $H(\mathbf{R}) = \|\mathbf{P} - \mathbf{R}\|_F^2 + \beta \mathrm{trace}(\mathbf{R}^T\mathbf{L}\mathbf{R})$, where $\mathbf{L} = \mathbf{L}^u + \mathbf{L}^b + \mathbf{L}^v + \mathbf{L}^h$, and then the objective function can be re-written as $J(\mathbf{R}) = H(\mathbf{R}) + \alpha\|\mathbf{R}\|_*$. Suppose $\mathbf{R}_{k-1}$ is the solution at the $(k-1)$-th step, we can update to $\mathbf{R}_k$ by solving the following optimization problem which quadratically approximates $J(\mathbf{R})$ by the second-order Taylor expansion of $H(\mathbf{R})$ at $\mathbf{R}_{k-1}$:

$$\begin{aligned} \mathbf{R}_k &= \arg\min_{\mathbf{R}} H(\mathbf{R}_{k-1}) + < \nabla H(\mathbf{R}_{k-1}), \mathbf{R} - \mathbf{R}_{k-1} > \\ &\quad + \frac{\delta}{2}\|\mathbf{R} - \mathbf{R}_{k-1}\|_F^2 + \alpha\|\mathbf{R}\|_* \\ &= \arg\min_{\mathbf{R}} \frac{\delta}{2}\|\mathbf{R} - \mathbf{G}_k\|_F^2 + \alpha\|\mathbf{R}\|_*, \end{aligned} \quad (8)$$

where the values of $\mathbf{G}_k$ and δ are defined as

$$\begin{aligned} \mathbf{G}_k &= \mathbf{R}_{k-1} - \frac{2}{\delta}\left(\mathbf{R}_{k-1} - \mathbf{P} + \alpha\mathbf{L}\mathbf{R}_{k-1}\right), \\ \delta &= 2\sigma_{max}\left(\mathbf{I} + \alpha\mathbf{L}\right), \end{aligned} \quad (9)$$

where δ satisfies the Lipschitz condition and $\sigma_{max}(\cdot)$ denotes the largest singular value. Note that the solution of Eq. (8) is $\mathbf{R}_k = \mathbf{U}\mathrm{diag}\lfloor\sigma - \frac{\alpha}{\delta}\rfloor_+\mathbf{V}^T$, where $\mathbf{U}\mathrm{diag}(\sigma - \frac{\alpha}{\delta})\mathbf{V}^T$ is the singular value of $\mathbf{G}_k$ [27]. Also, note that even with a large amount of images, the above optimization for profile refinement is tractable. To see this, for solving the singular values of $\mathbf{G}_k$, we can apply the trick to solve it by obtaining the singular values of $\mathbf{G}_k^T$, which can be much smaller. Meanwhile, for solving the largest singular values of $\mathbf{I} + \alpha\mathbf{L}$, we can adopt the efficient power method in matrix analysis.

After the above low-rank approximation, the average value of hierarchical representations of all the images after refinement $\mathbf{R}$ is seen as the user profile. The effectiveness of this profile refinement algorithm is shown in Figure 6. It shows that the proposed methods can refine the user profiles with the same user interest (*i.e.*, “bag” in this example) close to each other so that they have consistent user profile representations. It gives an intuitive interpretation of the expected better recommendation performance of the proposed profile refinement since it is often more accurate to calculate the user-user similarities in collaborative filtering approaches, and user-image similarities in content-based approaches.

5. EXPERIMENT

In this section, we systematically evaluate the effectiveness of our proposed profile learning, profile refinement algorithm and image recommendation using the profiles.

5.1 Data and Methodology

5.1.1 Experimental Data Set

We crawled the Pinterest data based on HTTP requests as there is no open API of Pinterest. The crawling protocol was as follows. First, we started from 50 popular pins from seed data in the fashion domain. Next, for each pinned image, we used a breadth-first search (BFS) strategy to crawl the boards which have pined or re-pined the image. The overall crawling process took a month. As a result, the data set consists of 1,239 users and 1,538,658 images.

In order to populate images into the constructed ontology in Section 3, we matched the items of user interest (*e.g.*, bundle names, tags) with entries in our constructed ontology to annotate the images of the corresponding concepts. Besides, the time when the user pinned the images can be obtained, therefore, we divided the images based on the pinning time for image recommendation. In order to train the profile ontology model, we split the data set equally into training/testing sets. Note that the testing set was used for testing ontology profile models, profile refinement algorithms and image recommendation. For profile refinement, the groundtruths were the same as the profile models. For recommendation, the average number of images per user were 433. In order to simulate real-world recommender system which has data from multiple domains, we added half the number of noisy data (*i.e.*, around 200) that are not in fashion domain.

5.1.2 Implementation Details

The underlying deep architecture we adopted in Section 3.2 is the deep convolutional neural network architecture proposed by Krizhevsky *et al.* [13]. Its inputs were the raw RGB pixel intensity values of a 224×224 image. Those values were forwarded through 5 convolutional neural layers (with pooling and ReLU non-linearities activation function along the way) and 3 fully-connected layers to determine its final neuron activities, namely, a distribution over the sibling concepts of user interest. As a result, neurons of the network in each layer were respectively 150,528-D, 290,400-D, 186,624-D, 64,896-D, 64,896-D, 43,264-D, 4,096-D, 4,096-D, and M-D, where M is the number of sibling concepts. We used the ImageNet pre-trained model, namely DeCAF [8], as the pre-trained network to initialize CNN and mtCNN. For visual features of the images used in Eq. (6) and other content-based baseline methods, we also adopted the 6-th layer output of DeCAF, which is a 4096-D feature vector.

For the sparse coding in section 3.1, we empirically initialized λ as 0.1. For multi-task convolutional neural network, we empirically set λ as 0.0001. For the profile refinement algorithm, we set $\alpha \in \{0.0001, 0.01, 0.1..., 10\}$, $\beta \in \{0.00001, 0.01, 0.1, ..., 10\}$, various pairs of (α, β) values were tried and the one with the best performance was chosen.

For all the experiments, we used an NVIDIA 780X GPU with 2304 cores, 3GB memory, and i7-2600 CPU with 3.40 GHz and 16G memory.

5.1.3 Compared Methods and Evaluation Metric

To evaluate the effectiveness of the proposed multi-task CNN (**mtCNN**), we compared it with state-of-art convolutional neural network [13]. For the model of each concept node in the ontology, we used the average precision (AP) as the evaluation metric.

To study the performance of our proposed profile refinement algorithm (**Ours**), two algorithms were employed as the baselines: a) **LR_ES_CC_TC** [33]: tag refinement algorithm low-rank, error sparsity, content consistency and tag correlation; and b) **TRVSC** [16]: tag refinement algorithm based on visual and semantic consistency. For evaluation metric, we used the widely used F_1-score.

To evaluate the effectiveness of user profiling methods, we proposed to use image recommendation based on user profiling. As mentioned in [3], current recommender systems generally fall into the following two categories: a) content-based recommendations, where users are recommended items similar to those they preferred in the past; and b) collaborative recommendations where users are recommended items that people with similar tastes and preferences liked in the past. Here, we extend the traditional recommendation methods using our proposed visual-based ontology profile. We used the content-based ontology profile vector to represent a user to calculate the user-item and user-user similarity. To evaluate the performance of visual-based profile in image recommendation, we compared it with the state-of-art methods: a) **CB**: this is the traditional content-based recommendation method [22], where users are recommended items similar to those they preferred in the past; b) **CF_WNMF** [10]: this method makes the use of non-negative matrix factorization on user and item graphs for collaborating Filtering; c) **CF_LDA** [26]: this method combines traditional collaborating methods with probabilistic topic modeling to provide an interpretable latent structure for users and items; d) **CB_UP**: this method extends the traditional content-based methods through representing the users with our profile ontology; and e) **CF_UP**: this method extends the traditional collaborating method by computing users’ similarities using our proposed profile ontology. We used mean AP (mAP) of the recommendation results as the evaluation metric.

5.2 Experimental Results

5.2.1 Evaluations of Profile Ontology Learning

Figure 7 illustrates the average precision values of different concept classifiers in the hierarchy. From this result, we can see that the multi-task convolutional neural network (**mtCNN**) at most levels achieves a mean average precision about 0.50, which is superior to traditional neural networks. These results demonstrate the effectiveness of **mtCNN** that makes use of hierarchical visual tasks. However, it can be seen that our proposed method has comparatively low performance on some concepts such as “zipper front dresses” and “skinny pants”. The reason for the low performance could be: (a) the distinctive attribute of those concepts is too fine-

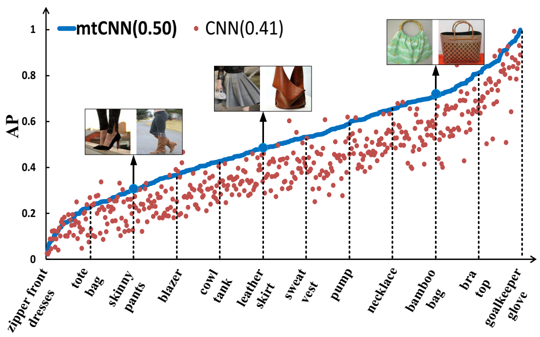


Figure 7: Performance of the 464 user profile models trained by CNN and the proposed mtCNN. mAPs are shown in brackets. Representative user interest and its two most confident images are shown.

grained, e.g. "zipper front", to differentiate those items correctly; and (b) these concepts tend to co-occur frequently with common items in an image. For example, the item "skinny pants" often co-occur with the item "high heels" in an image, then our method would recognize the "high heels" with "skinny pants" as "skinny pants". On the other hand, our method has quite good performance on other items such as "bamboo bag" and "goalkeeper glove". The reason for this is that those items have comparatively clean image samples.

5.2.2 Evaluations Of Profile Refinement

Figure 8 shows the detailed performance of image refinement for individual tags between our proposed approach and the baselines. From these results, we can see that the proposed method deploying social curation achieves an average F_1-score of 0.48, much higher than the other two methods. The superiority of our proposed profile refinement algorithms arises from two aspects: a) low rank; and b) multi-level social connections. Thanks to the content-centric network, multi-level connections are encoded to refine the user profiles. The experiment results explicitly illustrate that social curation services provide more structured and accurate information to infer user's preferences.

From Figure 8, we can observe that some classes may have comparatively lower F_1-score. For example, the F-score of the item "mini skirts" is about 0.25. It may be due to the noisy content-level connection since "mini skirts" and "mini dress" are often pinned into the same boards. Moreover, some cases fail because some images may not be that popular and therefore there exist only sparse and noisy content-level connections. In contrast, we observe that popular items tend to have higher F_1-scores;

Figure 9 shows some tag refinement results for some sample images produced by our approach. In the figure, the original user profiles come from tag annotation while the refined profiles comes from low-rank approximation. We can see that our approach can effectively correct and enrich the imprecise and incomplete image tags. For example, in Figure 9(c), our approach removes the irrelevant tags "dresses" and adds the more detailed tags "tote bag" and some other related tags such as "pink skirt" through social curation. Moreover, the enrichment capability of our proposed approaches can be seen in Figure 9(e) and (f), where there is only one irrelevant tag that shows people's intentions such as

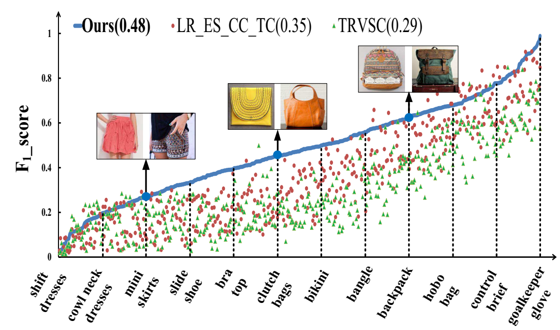


Figure 8: Performance of the three profile refinement methods. Mean F_1-scores are shown in brackets. The profile contains 464 user interests. Representative user interest in the profile and its two most confident images are shown.

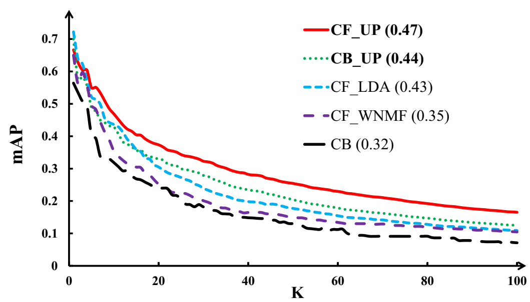


Figure 10: Performance (mAP@K) of the five recommendation methods. mAP@10 are shown in brackets.

"comfy" and "outfit" and after refining the incomplete tags by our approaches, the images are associated with reasonable tags. However, some cases fail because of lack of sibling samples. For example, the concepts "bow tie" and "ascot tie" under "tie" have fewer samples than others on the ontology.

5.2.3 Evaluations of Image Recommendation

Figure 10 shows the performance comparison between the proposed image recommendation methods and the other four state-of-the-art recommendation methods. We can observe that our proposed recommendation methods based on the visual-based ontology profile achieve the best performance in terms of MAP at all the top K results compared to the other methods. For example, our method improves the performance by 13.3%, 22.6%, 48.0%, 54.2% in terms of MAP at the top 20 results compared to the CB, CF_WNMF, CF_LDA and CB_UP respectively. This verifies the effectiveness of our proposed ontology profile in recommendation systems. However, if the user's interest is too general, the framework doesn't work well since our recommender system will recommend all the images. Some illustrative examples are shown in Figure 11.

The superiority of the proposed ontology profile arises from the following points: a) this profile models the user with a semantic hierarchy consisting of past users' interest; such hierarchy profile provide a more comprehensive interpretation of images of interest of users; and b) through

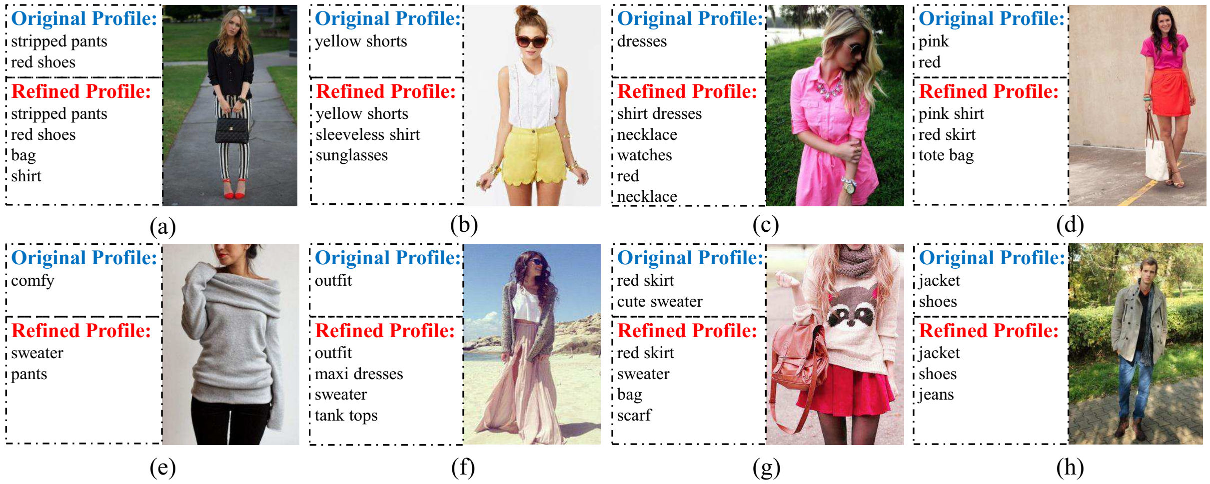


Figure 9: Illustrative profile refinement results by the proposed method.

computing the users' similarities based on this hierarchy profile, implicitly similar users are obtained which alleviates the sharing sparsity problem in traditional collaborative filtering recommendation methods. For example, there may exist many images that are rarely shared and would not be recommended. Besides, since our ontology construction does not totally rely on Wikipedia, but also the user comments which cover the real interest even if it falls into the long tail. Once a general domain is selected, our ontology can adapt to the true distribution of user interest.

6. CONCLUSIONS

In this paper, we targeted at content-based user profiling on the emerging social curation service (SCS). As compared to the conventional social network service (SNS), which focuses on user-user connections, SCS is based on the content-content connections curated by users. This is a unique characteristic of SCS, which inspires the idea of our profiling method. In particular, we investigated the fashion domain in the most popular SCS, namely Pinterest, on how social curation can help us tackle the existing difficulties in social media analysis. Specifically, we proposed to automatically construct a content-based user preference ontology and learn the ontological models to generate comprehensive user profiles. Then, we proposed to model the multi-level social relations offered by SCS to refine the user profiles in a low-rank recovery framework. Extensive experiments on 1,239 users and 1.5 million images are collected from Pinterest in fashion domain demonstrated the effectiveness of the proposed user profiling method, which outperforms the other state-of-the-art methods.

SCS is still in its infancy and thus requires a lot of explorations. For our initial exploration, we only focused on the fashion domain. However, in order to set up a real application running on a variety of SCSs, we should further investigate how to automatically determine the domains of users' interest. This is not a trivial task due to the extremely large diversity of user interests. Nevertheless, we believe that the proposed user profiling based on multimedia analysis with the help of social curation has a great potential to stimulate future research in social network.

7. ACKNOWLEDGEMENTS

This work was supported by the NUS-Tsinghua Extreme Search (NExT) project under the grant NO.: R-252-300-001-490 and the National Natural Science Foundation of China under the grant No.: 61303075.

8. REFERENCES

[1] F. Abel, Q. Gao, G.-J. Houben, and K. Tao. Analyzing user modeling on twitter for personalized news recommendations. In *UMAP*. 2011.

[2] F. Abel, E. Herder, G.-J. Houben, N. Henze, and D. Krause. Cross-system user modeling and personalization on the social web. *UMUAI*, 2013.

[3] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *TKDE*, 2005.

[4] Y. Cai and Q. Li. Personalized search by tag-based user profile and resource profile in collaborative tagging systems. In *ACM KDD*, 2010.

[5] M. Campbell. Online friendships light up shadow social networks. *NewScientist Magazine*, 2012.

[6] T.-S. Chua, H. Luan, M. Sun, and S. Yang. Next: Nus-tsinghua center for extreme search of user-generated content. *MultiMedia, IEEE*, 2012.

[7] M. Deshpande and G. Karypis. Item-based top-n recommendation algorithms. *ACM TOIS*, 2004.

[8] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. *arXiv preprint arXiv:1310.1531*, 2013.

[9] T. Evgeniou and M. Pontil. Regularized multi–task learning. In *ACM KDD*, 2004.

[10] Q. Gu, J. Zhou, and C. H. Ding. Collaborative filtering: Weighted nonnegative matrix factorization incorporating user and item graphs. In *SDM*, 2010.

[11] K. Ishiguro, A. Kimura, and K. Takeuchi. Towards automatic image understanding and mining via social curation. In *ICDM*, 2012.

[12] A. Kimura, K. Ishiguro, M. Yamada, A. Marcos Alvarez, K. Kataoka, and K. Murasaki.

Figure 11: Illustrative recommendation results from the proposed collaborative filtering based on our user profiles (CF_UP). Top nine recommended images for six users are shown.

Image context discovery from socially curated contents. In *ACM MM*, 2013.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

[14] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *TWEB*, 2007.

[15] J. Lin, K. Sugiyama, M.-Y. Kan, and T.-S. Chua. Addressing cold-start in app recommendation: latent user models constructed from twitter followers. In *SIGIR*, 2013.

[16] D. Liu, X.-S. Hua, M. Wang, and H.-J. Zhang. Image retagging. In *ACM MM*, 2010.

[17] M. Marszalek and C. Schmid. Semantic hierarchies for visual object recognition. In *CVPR*, 2007.

[18] J. McAuley and J. Leskovec. Image labeling on a network: using social-network metadata for image classification. In *ECCV*. 2012.

[19] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[20] S. Mittal, N. Gupta, P. Dewan, and P. Kumaraguru. The pin-bang theory: Discovering the pinterest world. *arXiv preprint arXiv:1307.4952*, 2013.

[21] R. Ottoni, J. P. Pesce, D. Las Casas, G. Franciscani Jr, W. Meira Jr, P. Kumaraguru, and V. Almeida. Ladies first: Analyzing gender roles and behaviors in pinterest. *ICWSM*, 2013.

[22] M. J. Pazzani and D. Billsus. Content-based recommendation systems. In *The adaptive web*. 2007.

[23] N. Qian. On the momentum term in gradient descent learning algorithms. *NN*, 1999.

[24] K.-C. Toh and S. Yun. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *PJO*, 2010.

[25] A. Ulges, M. Worring, and T. Breuel. Learning visual contexts for image annotation from flickr groups. *TMM*, 2011.

[26] C. Wang and D. M. Blei. Collaborative topic modeling for recommending scientific articles. In *ACM KDD*, 2011.

[27] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices by convex optimization. In *NIPS*, 2009.

[28] Z. Xu, Y. Zhang, Y. Wu, and Q. Yang. Modeling user posting behavior on social media. In *SIGIR*, 2012.

[29] S.-H. Yang, B. Long, A. Smola, N. Sadagopan, Z. Zheng, and H. Zha. Like like alike: Joint friendship and interest propagation in social networks. In *WWW*, 2011.

[30] H. Zhang, Y. Yang, H. Luan, S. Yan, and T.-S. Chua. Start from scratch: Towards automatically identifying, modeling, and naming visual attributes. In *ACM MM*, 2014.

[31] H. Zhang, Z.-J. Zha, Y. Yang, S. Yan, Y. Gao, and T.-S. Chua. Attribute-augmented semantic hierarchy: Towards bridging semantic gap and intention gap in image retrieval. In *ACM MM*, 2013.

[32] Y. Zhang, W. Chen, D. Wang, and Q. Yang. User-click modeling for understanding and predicting search-behavior. In *ACM KDD*, 2011.

[33] G. Zhu, S. Yan, and Y. Ma. Image tag refinement towards low-rank, content-tag prior and error sparsity. In *ACM MM*, 2010.

[34] I. Zukerman and D. W. Albrecht. Predictive statistical models for user modeling. *UMUAI*, 2001.