

Social Anchor-Unit Graph Regularized Tensor Completion for Large-Scale Image Retagging

Jinhui Tang¹, Xiangbo Shu¹, Zechao Li¹,
Yu-Gang Jiang¹, and Qi Tian², *Fellow, IEEE*

Abstract—Image retagging aims to improve the tag quality of social images by completing the missing tags, rectifying the noise-corrupted tags, and assigning new high-quality tags. Recent approaches simultaneously explore visual, user and tag information to improve the performance of image retagging by mining the tag-image-user associations. However, such methods will become computationally infeasible with the rapidly increasing number of images, tags and users. It has been proven that the anchor graph can significantly accelerate large-scale graph-based learning by exploring only a small number of anchor points. Inspired by this, we propose a novel Social anchor-Unit GrAph Regularized Tensor Completion (SUGAR-TC) method to efficiently refine the tags of social images, which is insensitive to the scale of data. First, we construct an anchor-unit graph across multiple domains (e.g., image and user domains) rather than traditional anchor graph in a single domain. Second, a tensor completion based on Social anchor-Unit GrAph Regularization (SUGAR) is implemented to refine the tags of the anchor images. Finally, we efficiently assign tags to non-anchor images by leveraging the relationship between the non-anchor units and the anchor units. Experimental results on a real-world social image database well demonstrate the effectiveness and efficiency of SUGAR-TC, outperforming the state-of-the-art methods.

Index Terms—Image retagging, anchor graph, tensor completion, image retrieval

1 INTRODUCTION

In the past decades, various social image retagging methods [1], [2], [3], [4], [5], [6] have been proposed to improve the tag quality of social images. At the early stage, image retagging methods refine tags of social images by utilizing semantic correlation among tags and visual similarity among images [7], [8]. A basic assumption of such methods is that two images with high visual similarity should have similar semantic tags, and vice versa. To well leverage the available image-tag inter-association, some researchers improved the matrix completion based on the low-rank decomposition to simultaneously recover the missing tags, and remove the noisy tags [9], [10]. Candes and Plan [11] have proven that the matrix completion model enables to complete the missing entries from a small number of observed entries in the original matrix between the dyadic data. However, the aforementioned methods achieve unsatisfied retagging results when there exists disambiguation between the visual content and label taxonomy (e.g., WordNet taxonomy [12]).

Intuitively, a set of images uploaded by one common user tends to have close relations. That is to say, the user information (e.g., user interests and backgrounds) bridges the inter-relationship

between tags and social images [13], especially for some certain tags, e.g., geo-related tags, event tags, etc. Therefore, several methods were proposed to simultaneously leverage the visual information, tag information, and user information for social image retagging. Sang et al. [2] attempted to model the inter-association among users, images and tags, and presented a tensor completion based on the low-rank decomposition to refine the social image tags. However, since the process of tensor completion generates many large-scale temporary matrices and tensors, it requires extremely large computational cost. To handle this problem, Tang et al. [4] proposed Tri-clustered Tensor Completion (TTC) to divide the original tag-image-user tensor among all tags, images, and users into several sub-tensors, and implement the sub-tensor completion in a parallel way. However, the computational cost still increases accompanied with the number of images increasing.

To this end, we aim to develop an efficient image retagging framework to assign high-quality tags to social images, regardless of the data scale. We investigate that anchor graph can significantly accelerate large-scale graph-based learning by only exploring a small number of anchor points on the graph [14], [15]. Inspired by this, we propose a Social anchor-Unit GrAph Regularization (SUGAR) model on a tag-image-user graph by exploring a small number of representative anchor units, where one anchor unit consists of an anchor image and an anchor user, as shown in Fig. 1. Accordingly, we propose a novel Social anchor-Unit GrAph Regularized Tensor Completion (SUGAR-TC) method to efficiently refine the tags of social images. The framework of SUGAR-TC is shown in Fig. 2, which mainly consists of three modules. (a) Anchor-unit graph construction module. We employ the co-clustering algorithm to obtain the representative anchor units in multiple domains (i.e., image and user domains) rather than the traditional anchors in a single domain, and then construct a social anchor-unit graph. (b) SUGAR tensor completion module. We propose a tensor completion model with Social anchor-Unit GrAph Regularization (SUGAR). It completes the anchor tensor by decomposing the non-anchor tensor to refine the tags of anchor images. (c) Anchor-aware tag assignment module. By leveraging the potential relationships between non-anchor units and anchor units, we use the weighted average tags of anchor images to efficiently assign high-quality tags to the non-anchor images.

Overall, the main contributions of this work are two-fold. (1) We propose an efficient Social anchor-Unit GrAph Regularized Tensor Completion (SUGAR-TC) method for social image retagging, even when the data scale is very large. (2) To the best of our knowledge, it is the first time that social anchor unit across multiple domains rather than the traditional anchors in a single domain is presented. In the experiments, the proposed SUGAR-TC gains superior performance on both effectiveness and efficiency compared with the state-of-the-art methods.

2 RELATED WORK

2.1 Social Image Retagging

The goal of social image retagging is to improve tag quality by recovering missing tags and removing noisy tags [16], [17]. It is an essential step for large-scale tag-based image retrieval. The image retagging task is closely related to tag refinement [4], [18], tag completion [3]. Overall, social image retagging methods have gone through two stages, i.e., image-tag association based image retagging [1], [3], [7], [8], [19], [20], [21], [22], [23], [24], [25], and tag-image-user association based image retagging [2], [4], [5], [26].

The image-tag association based methods focus on exploring inter- and intra- associations among images and tags. As one of the most classic works, Liu et al. [7] assumed that two images with

- J. Tang, X. Shu, and Z. Li are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China. E-mail: {jinhuitang, shuxb, zechao.li}@njjust.edu.cn.
- Y.-G. Jiang is with the School of Computer Science, Fudan University, Shanghai 201203, China. E-mail: ygj@fudan.edu.cn.
- Q. Tian is with the Noah's Ark Lab, Huawei, Shenzhen 518129, China. E-mail: tian.qi1@huawei.com.

Manuscript received 14 Apr. 2018; revised 27 Dec. 2018; accepted 15 Mar. 2019. Date of publication 24 Mar. 2019; date of current version 11 July 2019.

(Corresponding author: Xiangbo Shu.)

Recommended for acceptance by Steven C.H. Hoi.

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.

Digital Object Identifier no. 10.1109/TPAMI.2019.2906603

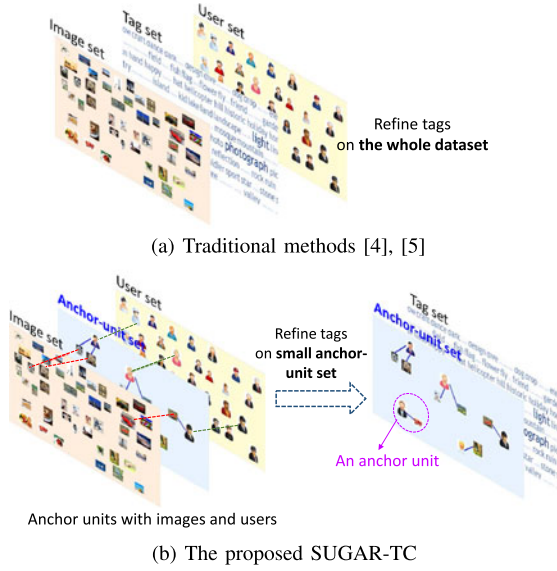


Fig. 1. Conventional methods refine tags of images by exploring the inter-association among all data. This comes at a price of increased computational cost when faced with a larger number of images. The proposed method refines tag on a smaller-scale anchor-unit set, and then assigns tags to all non-anchor images in an efficient way, regardless of the data scale.

higher visual similarity are more likely to have similar tags, or common tags. Yang et al. [8] proposed to mine the multi-tag intra-association for tag expansion and denoising. Chen et al. [20] proposed co-regularized learning with two classifiers by jointly mapping visual features and text features into a common subspace.

Matrix completion based on the low-rank decomposition is popularly applied for image retagging. As an early study, Zhu et al. [1] solved the image retagging problem by factorizing the image-tag inter-association matrix into an approximately low-rank completed matrix and a sparse error matrix, where this low-rank completed matrix reveals the image-tag associations. Besides, Feng et al. [27] theoretically analyzed that the matrix completion method is able to simultaneously recover missing tags and remove noisy tags even with a limited number of observations. Motivated by the success of low-rank model in image tagging, Xu et al. [9] proposed Non-linear Matrix Completion (NMC) for social image tagging by constructing the original image-tag matrix with a non-linear kernel mapping. Li et al. [10] divided the original image-tag matrix into several sub-matrices, and proposed a Locality Sensitive Low-Rank (LSLR) model on each sub-matrix to recover the image-tag matrix via matrix factorization.

Tag-image-user association based image retagging methods construct a tag-image-user inter-association tensor instead of the image-tag inter-association matrix to explore the important user information [2]. It can further improve the performance of image retagging with the help of user information. Tang et al. [4] proposed a Tri-clustered Tensor Completion (TTC) framework to first divide the original super-sparse tensor into several sub-tensors, and complete all sub-tensors regularized by a tensor kernel. However, to a certain degree, TTC will break the tag-image-user inter-association when dividing the original super-sparse tensor into several sub-tensors.

2.2 Anchor Graph-Based Learning

Graph-based learning methods have achieved impressive performance in various applications [28]. However, such methods require much computational cost when the size of data rapidly increases. Therefore, several strategies have been proposed to reduce the computational cost of graph-based learning. To sum up, these strategies can be summarized into three categories. The first strategy uses neighborhood propagation from the approximate neighborhood graph to current graph [29], [30], [31], [32]. The second strategy utilizes hashing to improve the performance in terms of efficiency [33], [34], [35], [36]. The last strategy employs the anchor graph to simultaneously reduce the computational cost and memory cost [15], [37], [38], [39], [40]. On an anchor graph, anchors covering vast data point cloud can predict the label for each data point, even when the data size is large. As a result, the anchor graph model has been successfully applied to solve many practical tasks, including image retrieval [41], face recognition [42], image classification [15], object tracking [43], and so on. In this work, we extend the anchor graph in a single domain to the anchor-unit graph across multiple domains, to simultaneously explore the association information among different domains.

3 THE PROPOSED FRAMEWORK

This section introduces the whole framework of SUGAR-TC, as shown in Fig. 2. It includes three main modules, i.e., Social anchor-unit graph construction module, SUGAR tensor completion module, and Anchor-aware tag assignment module. In this paper, tensors, matrices, vectors, variables and sets are denoted by calligraphic uppercase letters (e.g., \mathcal{A}), bold uppercase letters (e.g., \mathbf{T}), bold lowercase letters (e.g., \mathbf{d}), letters (e.g., m , l) and blackboard bold letters (e.g., \mathbb{I}), respectively. $\mathbf{T}_{i,j}$ denotes the i th row and j th column entry of \mathbf{T} . $\mathcal{T}_{i,j,k}$ denotes the i th row, j th column and k th tube entry of \mathcal{T} . $|\mathbb{I}|$ denotes the number of data in \mathbb{I} . $\mathcal{A}_{[1]}$, $\mathcal{A}_{[2]}$ and $\mathcal{A}_{[3]}$ are the matrix form of \mathcal{A} by accumulating the entries of \mathcal{A}

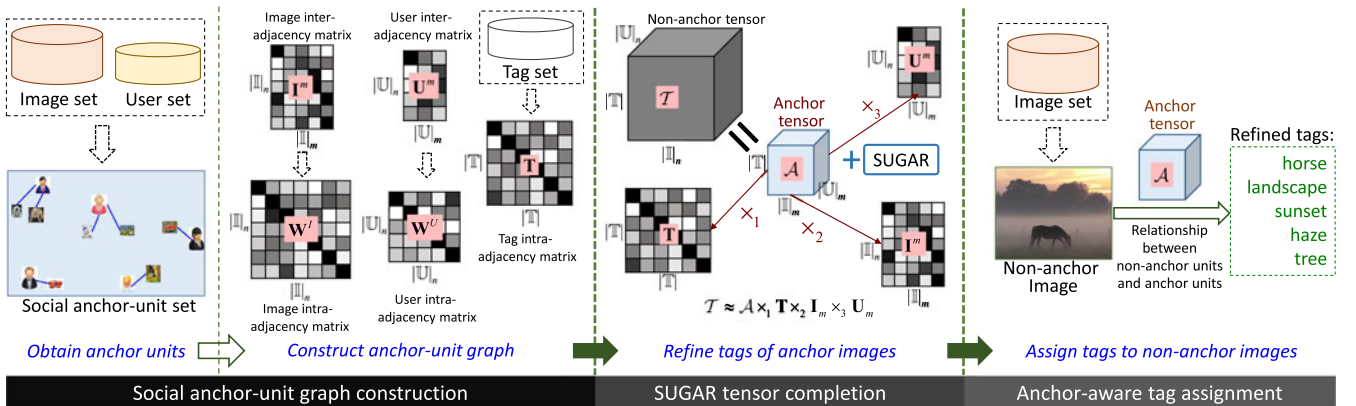


Fig. 2. The whole framework of SUGAR-TC for social image retagging. (a) Social anchor-unit graph construction: Obtain anchor units and construct anchor-unit graph with the inter- and intra- adjacency edges. (b) SUGAR tensor completion: refine tags of anchor images. (c) Anchor-aware tag assignment: assign tags to non-anchor images via the inter-association between non-anchor units and anchor units.

TABLE 1
Notations and Definitions

Notation	Definition
\mathbb{I}	Image set.
\mathbb{U}	User set.
\mathbb{T}	Tag set.
\mathbb{I}_n	Non-anchor image set.
\mathbb{U}_n	Non-anchor user set.
\mathbb{I}_m	Anchor image set.
\mathbb{U}_m	Anchor user set.
\mathcal{D}	Original tag-image-user tensor on \mathbb{T} , \mathbb{I} and \mathbb{U} .
\mathcal{T}	Original tag-image-user non-anchor tensor on \mathbb{T} , \mathbb{I}_n and \mathbb{U}_n .
\mathcal{A}_0	Original tag-image-user anchor tensor on \mathbb{T} , \mathbb{I}_m and \mathbb{U}_m .
\mathcal{A}	Refined tag-image-user anchor tensor. (<i>to be learned</i>)
\mathbf{I}_n	Inter-adjacency matrix between non-anchor image and anchor image.
\mathbf{U}_m	Inter-adjacency matrix between non-anchor user and anchor user.
\mathbf{W}_I	Intra-adjacency matrix between non-anchor image and non-anchor image.
\mathbf{W}_U	Intra-adjacency matrix between non-anchor user and non-anchor user.
\mathbf{T}	Intra-adjacency matrix between tag and tag.

along the row, column, and tube axes, respectively. For convenience, some important notations are defined in Table 1.

3.1 Background and Problem Definition

In this work, three types of data are collected from photo sharing websites, namely the image set $\mathbb{I} = \{x_i\}_{i=1}^{|\mathbb{I}|}$, the tag set $\mathbb{T} = \{t_j\}_{j=1}^{|\mathbb{T}|}$ and the user set $\mathbb{U} = \{u_k\}_{k=1}^{|\mathbb{U}|}$, where x_i , t_j and u_k denote the i th image, the j th tag, and the k th user, respectively. Since many tags of images freely given by users with various backgrounds and interests are ambiguous, noisy and incomplete, the original user-provided tags are weakly-supervised, which cannot represent the semantics of images well. If we directly implement the tag-based image retrieval on social images with original tags, the performance of retrieval will degrade. In this work, we aim to refine the tags of social images by utilizing the associations among tags, images and users. However, the numbers of images, tags and users are large, and even huge on photo sharing websites. This will dramatically increase the computational cost.

It has been proven that the anchor graph can significantly accelerate large-scale graph-based learning by exploring only a small number of anchor points. However, traditional anchor graph is constructed only in image domain. To leverage the user information, we extend the traditional anchor to anchor unit, which is a tuple including an anchor image and an anchor user. And the graph constructed on the anchor units is called anchor-unit graph.

Let \mathbb{I}_n , \mathbb{U}_n , \mathbb{I}_m , and \mathbb{U}_m indicate the non-anchor image set, the non-anchor user set, the anchor image set, and the anchor user set respectively, where $|\mathbb{I}_n| + |\mathbb{I}_m| = |\mathbb{I}|$ and $|\mathbb{U}_n| + |\mathbb{U}_m| = |\mathbb{U}|$, we construct a tag-image-user non-anchor tensor $\mathcal{T} \in \mathbb{R}^{|\mathbb{T}| \times |\mathbb{I}_n| \times |\mathbb{U}_n|}$ among all the tags, non-anchor images and non-anchor users, and a tag-image-user anchor tensor $\mathcal{A}_0 \in \mathbb{R}^{|\mathbb{T}| \times |\mathbb{I}_m| \times |\mathbb{U}_m|}$ among all the tags, anchor images and anchor users, in which some associations are incorrect or missing. Let $\mathcal{A} \in \mathbb{R}^{|\mathbb{T}| \times |\mathbb{I}_m| \times |\mathbb{U}_m|}$ denote the refined anchor tensor among the refined tags, anchor images and anchor users, this work first aims to learn \mathcal{A} with the help of the inter-association in \mathcal{T} and \mathcal{A}_0 . After that, benefiting from the tag-image-user association in \mathcal{A} , we can efficiently retag the non-anchor images based on the inter-relationship between non-anchor units and anchor units on the anchor-unit graph.

3.2 Social Anchor-Unit Graph Construction

To construct the anchor-unit graph, we need to obtain the anchor units from the image and user data. Generally speaking, the perfect

anchor units should satisfy two conditions: 1) they can adequately represent the distribution of image and user data; 2) the number of the anchor units should be much smaller than the number of images [37]. Existing methods employ the K-means clustering to obtain the traditional anchors [35], [38]. However, K-means algorithm performing in the single domain cannot cluster the image and user data simultaneously. Fortunately, co-clustering algorithm can construct the association between two types of data by an associated matrix, and then cluster rows and columns of this matrix simultaneously into several co-clusters [44]. Therefore, we can build an associated image-user matrix $\mathcal{D}_{[1]} \in \mathbb{R}^{|\mathbb{I}| \times |\mathbb{U}|}$ by accumulating entries of tag-image-user tensor \mathcal{D} along the tag axis, and adopt the co-clustering algorithm instead of the K-means algorithm to find C co-cluster centers. Subsequently, we select m_c image-user units (a user uploads this image) that is most close to the c th co-cluster center as the anchor unit. In total, we can obtain m image-user anchor units for C co-cluster centers, where $m = m_c \times C$.

Based on the m anchor units, we construct a social anchor-unit graph $\Omega\{\mathbb{I}_n, \mathbb{U}_n, \mathbb{I}_m, \mathbb{U}_m, \mathbb{T}, \omega\}$, where ω indicates the collection of the adjacency edges between the data points. To measure the weight of each edge in anchor-unit graph Ω , we design two types of inter-adjacency matrices between non-anchor data and anchor data (i.e., image inter-adjacency matrix and user inter-adjacency matrix), and three types of intra-adjacency matrices among non-anchor images, non-anchor users and tags (i.e., image intra-adjacency matrix, user intra-adjacency matrix, and tag intra-adjacency matrix).

Inter-adjacency matrices. First, we design the image inter-adjacency matrix $\mathbf{I}^m \in \mathbb{R}^{|\mathbb{I}_n| \times |\mathbb{I}_m|}$ between $|\mathbb{I}_n|$ non-anchor images and $|\mathbb{I}_m|$ anchor images, i.e.,

$$\mathbf{I}_{i,j}^m = \exp\left(-\frac{\|\mathbf{d}_{x_i} - \mathbf{d}_{x_j}\|_2^2}{\sigma^2}\right), \quad (1)$$

where σ is a parameter of the RBF kernel, \mathbf{d}_{x_i} and \mathbf{d}_{x_j} indicate features (i.e., CNN feature [45]) of a non-anchor image x_i and an anchor image x_j , respectively. Here, if $\mathbf{I}_{i,j}^m < 10^{-4}$, we set $\mathbf{I}_{i,j}^m = 0$ to obtain a sparse matrix \mathbf{I}^m for reducing the memory cost.

Second, we also design user inter-adjacency matrix $\mathbf{U}^m \in \mathbb{R}^{|\mathbb{U}_n| \times |\mathbb{U}_m|}$ between the $|\mathbb{U}_n|$ non-anchor users and $|\mathbb{U}_m|$ anchor users. Similar to [4], we assume that two users with higher co-occurrence are more likely to be related with each other, and vice versa, namely

$$\mathbf{U}_{i,j}^m = \frac{N(u_i, u_j)}{N(u_i) + N(u_j) - N(u_i, u_j)}, \quad (2)$$

where $N(u_i, u_j)$ denotes the number of groups that both a non-anchor user u_i and an anchor user u_j join, and $N(u_i)$ is the number of groups that a non-anchor user u_i joins, $N(u_j)$ is the number of groups that an anchor user u_j joins.

Intra-Adjacency Matrices. We first design image intra-adjacency matrix $\mathbf{W}^I \in \mathbb{R}^{|\mathbb{I}_n| \times |\mathbb{I}_n|}$ among the non-anchor images. One strategy is that we can measure the association between two non-anchor images by Eq. (1). However, the number of non-anchor images is much larger than the number of anchor images, which will increase the computational cost. Alternately, since anchor images are very representative in the whole image set, we can measure the association between non-anchor images by exploring links among common anchor images, which has also been claimed in [14], [15]. For two non-anchor images, the more number of common anchor images the two non-anchor images share, the stronger association such two non-anchor images have. Thus, $\mathbf{W}^I \in \mathbb{R}^{|\mathbb{I}_n| \times |\mathbb{I}_n|}$ can be computed as follows,

$$\mathbf{W}^I = \mathbf{I}^m (\mathbf{A}^I)^{-1} (\mathbf{I}^m)^T, \quad (3)$$

where the diagonal matrix Λ^I is defined as $\Lambda_{j,j}^I = \sum_{i=1}^{|\mathbb{I}_n|} \mathbf{I}_{i,j}^m$ ($1 \leq j \leq |\mathbb{I}_m|$). From Eq. (3), if $\mathbf{W}_{i,j}^I > 0$, it means two non-anchor images share at least one anchor user, and otherwise $\mathbf{W}_{i,j}^I = 0$.

Second, we use the same idea to design user intra-adjacency matrix $\mathbf{W}^U \in \mathbb{R}^{|\mathbb{U}_n| \times |\mathbb{U}_n|}$ among the non-anchor users, as follows,

$$\mathbf{W}^U = \mathbf{U}^m (\Lambda^U)^{-1} (\mathbf{U}^m)^T, \quad (4)$$

where the diagonal matrix Λ^U is defined as $\Lambda_{j,j}^U = \sum_{i=1}^{|\mathbb{U}_n|} \mathbf{U}_{i,j}^m$.

Following the definitions in [4], [5], we design the tag intra-adjacency matrix $\mathbf{T} \in \mathbb{R}^{|\mathbb{T}| \times |\mathbb{T}|}$ by using the categorical relations and the co-occurrence, i.e.

$$\mathbf{T}_{i,j} = a_1 \frac{N(t_i, t_j)}{N(t_i) + N(t_j) - N(t_i, t_j)} + a_2 \frac{2 \cdot C(L(t_i, t_j))}{C(t_i) + C(t_j)}, \quad (5)$$

where a_1 and a_2 denote the weight coefficients ($a_1 + a_2 = 1$); $N(t_i)$ denotes the occurrence count of tag t_i in a dataset; $N(t_i, t_j)$ denotes the co-occurrence count for tags t_i and t_j in a dataset; $C(t_i) = -\log(p(t_i))$ is the information content of tag t_i ; $p(t_i)$ is the probability of tag t_i ; and $L(t_i, t_j)$ is the least common subsumer of tags t_i and t_j in the WordNet taxonomy [12]. The least common subsumer of two tags in WordNet is the sumer that does not have any children that are also the subsumer of two tags.

3.3 SUGAR Tensor Completion

It has been proved that anchor graph can efficiently deal with the standard multi-class semi-supervised learning problem [37]. Motivated by this, we extend the existing Anchor Graph Regularization [38] to a novel Social anchor-Unit GrAph Regularization (SUGAR). Specifically, it can be assumed that two images linked by the same anchor image should have closely similar tags and closely related users, two users linked by the same anchor user prefer to upload images with closely similar tags. Then, we can define the regularization as follows,

$$\begin{aligned} \Theta_1 = & \frac{\lambda_1}{2} \sum_{i=1, j=1}^{|\mathbb{I}_n|} \mathbf{W}_{i,j}^I \|\mathcal{A} \times_2 \mathbf{I}_{i,:}^m - \mathcal{A} \times_2 \mathbf{I}_{j,:}^m\|_F^2 \\ & + \frac{\lambda_2}{2} \sum_{i=1, j=1}^{|\mathbb{U}_n|} \mathbf{W}_{i,j}^U \|\mathcal{A} \times_3 \mathbf{U}_{i,:}^m - \mathcal{A} \times_3 \mathbf{U}_{j,:}^m\|_F^2, \end{aligned} \quad (6)$$

where λ_1 and λ_2 are nonnegative coefficients to control the penalty of the corresponding regularization. The matrix $\mathcal{A} \times_2 \mathbf{I}_{i,:}^m$ denotes the 2-mode product¹ of \mathcal{A} and $\mathbf{I}_{i,:}^m$, and describes the inter-association between the tags and the anchor users corresponding to the i th non-anchor image. The matrix $\mathcal{A} \times_3 \mathbf{U}_{i,:}^m$ describes the inter-association between the tags and the anchor images corresponding to the i th non-anchor user.

The great number of original tags provided by users on web-sites involve the important supervised information, which can guide us to mine the inter-association among images, tags, and users [2]. Accordingly, we construct the tag-image-user non-anchor tensor $\mathcal{T} \in \mathbb{R}^{|\mathbb{T}| \times |\mathbb{I}_n| \times |\mathbb{U}_n|}$ among tags, non-anchor images and non-anchor users. Specifically, if the i th non-anchor image uploaded by the k th non-anchor user is annotated with the j th tag, we set $\mathcal{T}_{i,j,k} = 1$, otherwise $\mathcal{T}_{i,j,k} = 0$, where $1 \leq i \leq |\mathbb{T}|$, $1 \leq j \leq |\mathbb{I}_n|$, and $1 \leq k \leq |\mathbb{U}_n|$. For a tensor with a few available entries, the tensor completion algorithm can estimate missing entries and remove the noisy ones by reconstructing an approximately low-rank tensor $\tilde{\mathcal{T}}$, as follows,

$$\min_{\tilde{\mathcal{T}}} \|\mathcal{T} - \tilde{\mathcal{T}}\|_F^2. \quad (7)$$

The Tucker decomposition [46] of tensor provides a factorization way to solve the low-rank tensor $\tilde{\mathcal{T}}$ by the following objective function, i.e.,

$$\min_{\tilde{\mathcal{T}}} \|\mathcal{T} - \mathcal{S} \times_1 \mathbf{B} \times_2 \mathbf{C} \times_3 \mathbf{D}\|_F^2, \quad (8)$$

where $\tilde{\mathcal{T}} = \mathcal{S} \times_1 \mathbf{B} \times_2 \mathbf{C} \times_3 \mathbf{D}$. Here, \mathcal{S} denotes the core tensor, \mathbf{B} , \mathbf{C} , and \mathbf{D} denote the factor matrices. Since \mathbf{T} , \mathbf{I}^m and \mathbf{U}^m describe the associations between tags and tags, non-anchor images and anchor images, as well as non-anchor users and anchor users, respectively, by setting $\mathbf{B} = \mathbf{T}$, $\mathbf{C} = \mathbf{I}^m$ and $\mathbf{D} = \mathbf{U}^m$, the learned core tensor \mathcal{S} can be regarded as the refined tensor \mathcal{A} . To well leverage the weakly supervised tag-image-user association, we introduce a new regularization term $\|\mathcal{A} - \mathcal{A}_0\|_F^2$, which constrains the learned \mathcal{A} to be consistent with the original tag-image-user anchor tensor \mathcal{A}_0 . Finally, we can obtain the refined tensor \mathcal{A} by minimizing the following tensor completion function Θ_2 , i.e.,

$$\Theta_2 = \|\mathcal{T} - \mathcal{A} \times_1 \mathbf{T} \times_2 \mathbf{I}^m \times_3 \mathbf{U}^m\|_F^2 + \alpha \|\mathcal{A} - \mathcal{A}_0\|_F^2 + \beta \|\mathcal{A}\|_F^2, \quad (9)$$

where α and β are nonnegative parameters to control the penalty of the regularization. In Eq. (9), the learned \mathcal{A} is a low-rank and compact tensor, which can reveal the association among anchor-images, anchor-users and tags.

We integrate social anchor-graph regularization (i.e., Eq. (6)) into Eq. (9), and then obtain the completed anchor tensor \mathcal{A} by minimizing an objective function $\Theta = \Theta_1 + \Theta_2$, as follows,

$$\begin{aligned} \min_{\mathcal{A}} \Theta = & \min_{\mathcal{A}} \|\mathcal{T} - \mathcal{A} \times_1 \mathbf{T} \times_2 \mathbf{I}^m \times_3 \mathbf{U}^m\|_F^2 + \alpha \|\mathcal{A} - \mathcal{A}_0\|_F^2 + \beta \|\mathcal{A}\|_F^2 \\ & + \frac{\lambda_1}{2} \sum_{i=1, j=1}^{|\mathbb{I}_n|} \mathbf{W}_{i,j}^I \|\mathcal{A} \times_2 \mathbf{I}_{i,:}^m - \mathcal{A} \times_2 \mathbf{I}_{j,:}^m\|_F^2 \\ & + \frac{\lambda_2}{2} \sum_{i=1, j=1}^{|\mathbb{U}_n|} \mathbf{W}_{i,j}^U \|\mathcal{A} \times_3 \mathbf{U}_{i,:}^m - \mathcal{A} \times_3 \mathbf{U}_{j,:}^m\|_F^2. \end{aligned} \quad (10)$$

In this work, we learn \mathcal{A} in an iteratively updating way. Specifically, we compute the partial derivatives of the objective function Θ with respect to \mathcal{A} , i.e.,

$$\begin{aligned} \frac{\partial \Theta}{\partial \mathcal{A}} = & 2\mathcal{A} \times_1 (\mathbf{T}^T \mathbf{T}) \times_2 ((\mathbf{I}^m)^T \mathbf{I}^m) \times_3 ((\mathbf{U}^m)^T \mathbf{U}^m) \\ & - 2\mathcal{T} \times_1 \mathbf{T}^T \times_2 (\mathbf{I}^m)^T \times_3 (\mathbf{U}^m)^T + 2\alpha(\mathcal{A} - \mathcal{A}_0) + 2\beta\mathcal{A} \\ & + \lambda_1 \sum_{i=1, j=1}^{|\mathbb{I}_n|} \mathbf{W}_{i,j}^I (\mathcal{A} \times_2 (\mathbf{a}_i^T \mathbf{a}_i) - \mathcal{A} \times_2 (\mathbf{a}_i^T \mathbf{a}_j)) \\ & + \lambda_1 \sum_{i=1, j=1}^{|\mathbb{I}_n|} \mathbf{W}_{i,j}^I (\mathcal{A} \times_2 (\mathbf{a}_j^T \mathbf{a}_j) - \mathcal{A} \times_2 (\mathbf{a}_j^T \mathbf{a}_i)) \\ & + \lambda_2 \sum_{i=1, j=1}^{|\mathbb{U}_n|} \mathbf{W}_{i,j}^U (\mathcal{A} \times_3 (\mathbf{b}_i^T \mathbf{b}_i) - \mathcal{A} \times_3 (\mathbf{b}_i^T \mathbf{b}_j)) \\ & + \lambda_2 \sum_{i=1, j=1}^{|\mathbb{U}_n|} \mathbf{W}_{i,j}^U (\mathcal{A} \times_3 (\mathbf{b}_j^T \mathbf{b}_j) - \mathcal{A} \times_3 (\mathbf{b}_j^T \mathbf{b}_i)) \\ = & 2\mathcal{A} \times_1 (\mathbf{T}^T \mathbf{T}) \times_2 ((\mathbf{I}^m)^T \mathbf{I}^m) \times_3 ((\mathbf{U}^m)^T \mathbf{U}^m) \\ & - 2\mathcal{T} \times_1 \mathbf{T}^T \times_2 (\mathbf{I}^m)^T \times_3 (\mathbf{U}^m)^T + 2\alpha(\mathcal{A} - \mathcal{A}_0) + 2\beta\mathcal{A} \\ & + 2\lambda_1 \sum_{i=1, j=1}^{|\mathbb{I}_n|} \mathbf{W}_{i,j}^I (\mathcal{A} \times_2 (\mathbf{a}_i^T \mathbf{a}_i) - \mathcal{A} \times_2 (\mathbf{a}_i^T \mathbf{a}_j)) \\ & + 2\lambda_2 \sum_{i=1, j=1}^{|\mathbb{U}_n|} \mathbf{W}_{i,j}^U (\mathcal{A} \times_3 (\mathbf{b}_i^T \mathbf{b}_i) - \mathcal{A} \times_3 (\mathbf{b}_i^T \mathbf{b}_j)), \end{aligned} \quad (11)$$

1. More details about n -mode product can be found in [46].

TABLE 2
Average F-Scores, Precisions, and Recalls Obtained by Different Methods for Image Retagging

Methods	OT	TRVSC [7]	LR [1]	NMC [9]	LSLR [10]	MRTF [2]	TTC1 [4]	TTC2 [4]	SUGAR-TC
Average F-scores	0.379	0.390	0.410	0.417	0.435	0.423	0.447	0.458	0.485
Average Precision	0.452	0.473	0.532	0.546	0.549	0.554	0.580	0.581	0.586
Average Recall	0.326	0.331	0.333	0.337	0.360	0.342	0.364	0.378	0.414

where $\mathbf{a}_i = \mathbf{I}_{i,:}^m$, $\mathbf{a}_j = \mathbf{I}_{j,:}^m$, $\mathbf{b}_i = \mathbf{U}_{i,:}^m$, and $\mathbf{b}_j = \mathbf{U}_{j,:}^m$. Therefore, the multiplicative updating rule [47] of \mathcal{A} is

$$\mathcal{A}_{i,j,k} = \mathcal{A}_{i,j,k} \frac{(\mathcal{H} + \alpha \mathcal{A}_0 + \lambda_1 \mathcal{Q} + \lambda_2 \mathcal{P})_{i,j,k}}{(\mathcal{G} + (\alpha + \beta) \mathcal{A} + \lambda_1 \mathcal{U} + \lambda_2 \mathcal{V})_{i,j,k}}, \quad (12)$$

where

$$\begin{aligned} \mathcal{H} &= \mathcal{T} \times_1 \mathbf{T}^T \times_2 (\mathbf{I}^m)^T \times_3 (\mathbf{U}^m)^T, \\ \mathcal{G} &= \mathcal{A} \times_1 (\mathbf{T}^T \mathbf{T}) \times_2 ((\mathbf{I}^m)^T \mathbf{I}^m) \times_3 ((\mathbf{U}^m)^T \mathbf{U}^m), \\ \mathcal{Q} &= \sum_{i=1, j=1}^{|\mathcal{I}_n|} \mathbf{W}_{i,j}^I (\mathcal{A} \times_2 (\mathbf{a}_i^T \mathbf{a}_j)), \\ \mathcal{P} &= \sum_{i=1, j=1}^{|\mathcal{U}_n|} \mathbf{W}_{i,j}^U (\mathcal{A} \times_3 (\mathbf{b}_i^T \mathbf{b}_j)), \\ \mathcal{U} &= \sum_{i=1, j=1}^{|\mathcal{I}_n|} \mathbf{W}_{i,j}^I (\mathcal{A} \times_2 (\mathbf{a}_i^T \mathbf{a}_i)), \\ \mathcal{V} &= \sum_{i=1, j=1}^{|\mathcal{U}_n|} \mathbf{W}_{i,j}^U (\mathcal{A} \times_3 (\mathbf{b}_i^T \mathbf{b}_i)). \end{aligned} \quad (13)$$

The details of algorithm are described in Algorithm 1. Here, \mathcal{A} is initialized by $\mathcal{A}_0 + \mathcal{E}$, where \mathcal{E} is a random small-disturbed tensor with mean 0. The convergence criterion is that the iterations stop when the relative cost of the objective function is smaller than a predefined threshold 10^{-5} . The proposed SUGAR-TC achieves convergence after about 450 iterations in the experiments.

Algorithm 1. SUGAR Tensor Completion

Input:

Original non-anchor tensor \mathcal{T} , original anchor tensor \mathcal{A}_0 on the anchor-unit set, adjacency matrices \mathbf{T} , \mathbf{I}^m , \mathbf{U}^m , \mathbf{W}^I and \mathbf{W}^U , parameters α , β , λ_1 and λ_2 .

Output:

Completed anchor tensor \mathcal{A} .

Initialization: \mathcal{A} .

- 1: **repeat**
- 2: Update \mathcal{A} by Eq. (12).
- 3: **until** Convergence.
- 4: **return** \mathcal{A} .

3.4 Anchor-Aware Tag Assignment

After obtaining the completed tensor \mathcal{A} , we accumulate its entries along the user axis and image axis to acquire the desired tag-image association matrix $\mathcal{A}_{[3]}$ and tag-user association matrix $\mathcal{A}_{[2]}$, respectively. Then, we can employ the completed tags associated with anchor units to predict tags for non-anchor images. For one non-anchor image \mathbf{x}_i ($i = 1, 2, \dots, |\mathcal{I}_n|$) with an available user \mathbf{u}_k ($k = 1, 2, \dots, |\mathcal{U}_n|$), we use the weighted average of tags of s nearest-neighbor anchor units to estimate its tag vector \mathbf{y}_i , as follows,

$$\mathbf{y}_i = \frac{\gamma \mathbf{I}_{i, < i >}^m \left[(\mathbf{A}_{[3]})_{:, < i >} \right]^T + (1 - \gamma) \mathbf{U}_{k, < k >}^m \left[(\mathbf{A}_{[2]})_{:, < k >} \right]^T}{s}, \quad (14)$$

where $< i >$ is an index set of s closest anchor-images of \mathbf{x}_i , $< k >$ is an index set of users corresponding to these s closest anchor-images, γ is a parameter to control the degrees of tag-image and tag-user associations. Finally, we rank the elements of \mathbf{y}_i based on

the values in descending order, and select top-10 tags as the final tags of \mathbf{x}_i .

4 EXPERIMENTS

4.1 Experimental Settings

We conduct experiments on a real-world social image dataset NUS-WIDE-128 [48] to evaluate the performance of the proposed SUGAR-TC method. It is extended from the widely-used NUS-WIDE dataset [49], and contains 269,648 images, 5,018 user-provides tags crawled from the Internet, as well as the manually labeled ground-truth of 128 predefined concepts for evaluation. In this dataset, some user IDs are invalid or unavailable, thus we delete the corresponding images. Finally, we obtain 247,849 images with 49,528 user IDs and the information of user groups.

In the experiments, we compare the proposed method with TRVSC [7], LR [1], LSLR [10], NMC [9], MRTF [2], TTC1 [4] and TTC2 [4]. Besides, we also set a baseline method called Original Tagging (OT), where the labels used for evaluation are the user-provided tags crawled from Internet. To evaluate the performance, the widely-used F-score is adopted. For each concept, F-score is calculated as $\text{F-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$. And then the average F-score over all the concepts is reported. For fair comparison, we use the CNN features [45] for all methods, and all methods in the experiments are implemented on a server with an 8-core 2.67 GHz CPU and 32 GB memory.

4.2 Results and Analysis

In this section, we conduct experiments to evaluate the effectiveness of the proposed SUGAR-TC, and present the results of the average F-score, average precision, and average recall obtained by all methods in Table 2, respectively. It can be observed that the proposed SUGAR-TC achieves the best results for image retagging compared with the related methods. MRTF, TTC1 and TTC2 utilize the inter- and intra- associations among images, tags and users, thus they perform better than TRVSC, LR, NMC and LSLR in terms of F-score. In particular, since TTC1 and TTC2 address the super-sparse problem existing in the original tensor, they gain better performance than MRTF that directly completes the entries in the original tensor. However, to a certain degree, both TTC1 and TTC2 break the latent inter-association among images, tags and users, when dividing the original tensor into several sub-tensors. In turn, the proposed SUGAR-TC did not break the inherent tag-image-user inter-association. Thus, the F-score obtained by SUGAR-TC is 0.485, which is about 0.027 higher than 0.458 of TTC2.

We further detail the F-scores obtained by different methods on all the 128 predefined concepts, as shown in Fig. 3. We can see that SUGAR-TC achieves the highest F-scores on most of these concepts. We can also see that all image retagging methods improve the quality of almost all concepts compared with OT. By adding user information, MRTF, TTC1, TTC2 and SUGAR-TC perform better than LR, NMC, LSLR and TRVSC, especially for some summarized or complex tags (e.g., “military”, “nighttime”, and “cityscape”). In particular, for the frequently-used tags (e.g., “bus”, “flowers”, and “sunset”, etc) in Fig. 3, SUGAR-TC shows remarkable improvement than TTC1 and TTC2, one reason may be that TTC1 and TTC2 break



Fig. 3. The comparisons of detailed F-scores of different methods. For those concepts denoted with the black bounding box, the proposed SUGAR-TC shows remarkable improvements. Best view in color.

the tag-image-user inter-association for such common tags when decomposing the original tensor into several sub-tensors.

We also show some tag assignment results for some specific cases, as shown in Fig. 4. We can see that SUGAR-TC can effectively assign tags of images, even though there are some globally complex images (e.g., Fig. 4c) and locally abstract images (e.g., Fig. 4d), since SUGAR-TC can simultaneously mine the context information between users and images, and the semantic relation between images and tags. Besides, although some tags, i.e., geo-related tags, event tags, and time-related tags, are hardly inferred by only using visual information, SUGAR-TC considering extra user information can correctly infer them. For example, as shown in Fig. 4a, the train is a landmark of Japan, thus this image should be tagged with the geo-related tags “Japan”. It is very difficult to reveal the relation between the tag “Japan” and this image by only mining the visual and tag information, but SUGAR-TC can do it well based on the user background information. Similarly in Fig. 4b, this image shows the Obamas speaking of the United States presidential election, it is reasonable to assign the event tag “election” to this image.

4.3 Computational Cost Analysis

To illustrate the efficiency of the proposed SUGAR-TC, we also compare the computational cost of SUGAR-TC and the closely

related methods, including MRTF [2], TTC1 [4], and TTC2 [4]. Table 3 lists the computing time and memory cost of these methods in details. We can see that: 1) SUGAR-TC with total computing time 4.2h runs much faster than MRTF, TTC1 and TTC2; 2) the memory cost of SUGAR-TC is more than TTC1 and TTC2, but less than

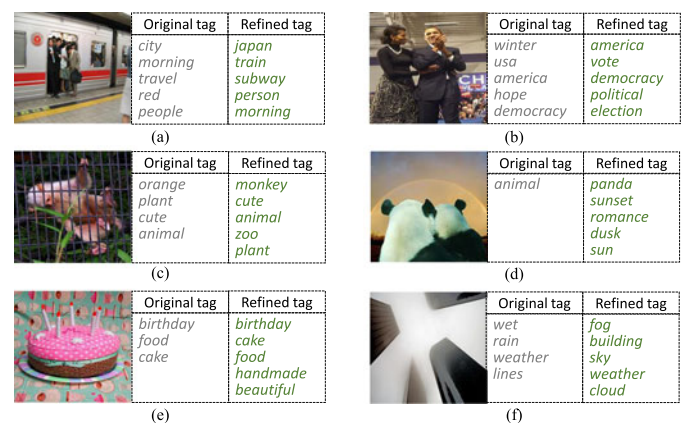


Fig. 4. Some results of image retagging obtained by SUGAR-TC.

TABLE 3
Detailed Comparisons of Computational Time (Hour)
and Memory (GB) on NUS-WIDE-128

Method	MRTF [2]	TTC1 [4]	TTC2 [4]	SUGAR-TC
Clustering	-	4.3/8.6	4.3/8.6	2.1/5.5
Adjacency matrix	1.6/1.8	1.6/1.8	1.6/1.8	0.7/0.07
Tensor completion	12.3/10.3	5.7/6.6	29.6/8.2	1.2/10.1
Tag assignment	-	-	-	0.2/0.08
Total time	13.9	11.6	35.5	4.2

" / " Denotes "Time/Memory".

MRTF. By jointly considering the accuracy, computing time and memory cost, SUGAR-TC is more applicable for image retagging.

4.4 Parameter Sensitivity Analysis

In the proposed SUGAR-TC method, there are several hyper-parameters to be set in advance. Following [4], the weight coefficients a_1 and a_2 in Eq. (5) are set to 0.9 and 0.1 respectively. The radius parameter δ in Eq. (1) is set to 2.5. The number of anchor units is set as $m_c = 10$ for each co-cluster center. In the following, we will discuss the sensitiveness of other parameters. Generally, for the tag refinement task, most of methods use a subset of the full dataset as the validation set to tune the hyper-parameters of the learning model. This subset is obtained by randomly selecting about 1/5 or 1/10 of images from the full dataset [2], [4], [10]. In our experiments, we use a subset including 25000 images, i.e., about 1/10 of NUS-WIDE-128, to tune the hyper-parameters of all the compared methods by using the grid search strategy. Subsequently, we set the tuned hyper-parameters to the learning model. For the image retagging task, all images with the noise-corrupted tags and users in NUW-WIDE-128 are used to learn the model, and then the learned model completes the missing tags and rectifies the noise-corrupted tags of all images.

For the number of image centers (denoted by C_i) and the number of user clusters (denoted by C_u), we tune them by $C_i \in \{10, 20, 40, 60, 80, 100, 200, 500\}$ and $C_u \in \{3, 6, 12, 18, 24, 30, 50, 200\}$, respectively. The corresponding F-scores are shown in Figs. 5a and 5b, respectively. From the results, we can see that the best result is obtained with $C_i = 40$ and $C_u = 12$. When these two parameters are set within $[20, 80]$ and $[6, 18]$ respectively, the performance changes slightly. In the experiments, the default number of image clusters and user clusters are set to 40 and 12, respectively.

For α , β , λ_1 , λ_2 , and γ , we tune them by $\alpha = \{0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1\}$, $\beta = \{0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1\}$, $\lambda_1 = \{0, 0.01, 0.05, 0.1, 0.25, 0.75, 1, 10\}$, $\lambda_2 = \{0, 0.01, 0.05, 0.1, 0.25, 0.75, 1, 10\}$, and $\gamma = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$. The corresponding results are shown in Figs. 5c, 5d, and 5e. SUGAR-TC achieves the best result with $\alpha = 0.005$, $\beta = 0.001$, $\lambda_1 = 0.1$, $\lambda_2 = 0.05$, and $\gamma = 0.8$, which are set as the default values of these parameters in the experiments.

4.5 Application: Image Retrieval

In this work, for each image, image retagging task can not only add and remove tags, but also re-rank the completed tag list by assigning

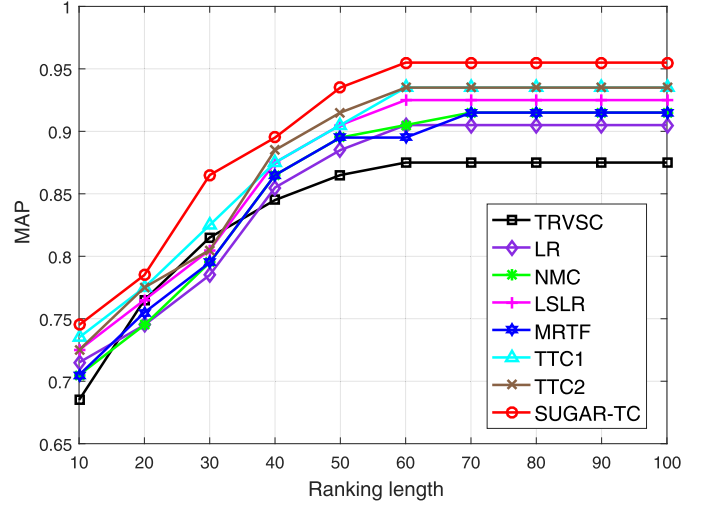


Fig. 6. MAP of different methods for tag-based image retrieval.

tags with different confidence scores. For each image, we can obtain a tag ranking list, which can improve the performance of tag-based image retrieval. Therefore, we conduct experiment of tag-based image retrieval to illustrate the effectiveness of SUGAR by comparing with other related methods, including TRVSC [7], LR [1], NMC [9], LSLR [10], MRTF [2], TTC1 [4] and TTC2 [4].

Following the experimental setting in [7], we perform tag-based social image retrieval with ten queries on NUS-WIDE-128, i.e., *birds, building, butterfly, dog, fish, flowers, horses, plane, sunset* and *zoo*. Average Precision (AP) is used as the evaluation metric. Specifically, given a ranked list with length T , AP is defined as $AP = \frac{1}{R} \sum_{r=1}^T p(r) \cdot \chi(r)$. Here, R is the number of relevant images in this ranked list, $p(r)$ is the precision at cut-off top r , $\chi(r)$ is an indicator function: $\chi(r) = 1$ if the r th image is relevant with respect to the ground-truth concept, otherwise $\chi(r) = 0$. Finally, we use MAP over all queries to evaluate the overall performance. Fig. 6 shows MAP values with different ranking length T . We can see that the proposed SUGAR-TC outperforms the other four image retagging methods. This demonstrates SUGAR-TC is more effective than other related image retagging methods in terms of tag-based social image retrieval.

5 CONCLUSIONS

In this work, we propose a novel Social anchor-Unit GrAph Regularized Tensor Completion (SUGAR-TC) method to efficiently refine the tags of social images, regardless of the data scale. First, we utilize the co-clustering algorithm to obtain the representative anchor units (anchor image and anchor user) across image and user domains rather than traditional anchors in a single domain, and then construct an anchor-unit graph with multiple intra/inter adjacency edges. Second, we present a SUGAR tensor completion to refine tags of anchor images. Finally, we efficiently assign high-quality tags to all non-anchor images by leveraging the potential relationships between non-anchor units and anchor units. Experimental

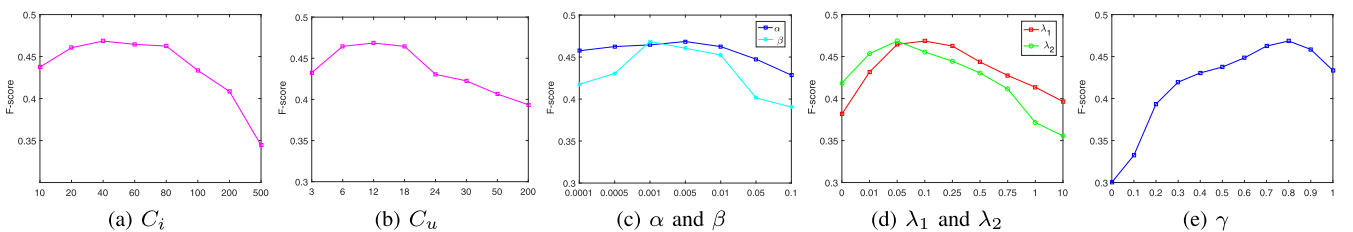


Fig. 5. The curves of F-scores obtained by varying one parameter with the other parameters fixed to be the default values.

results on a real-world social image database demonstrate the effectiveness and efficiency of SUGAR-TC compared with the state-of-the-art methods.

ACKNOWLEDGMENTS

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB1001001, in part by the National Natural Science Foundation of China under Grant 61732007 and Grant 61702265.

REFERENCES

- [1] G. Zhu, S. Yan, and Y. Ma, "Image tag refinement towards low-rank, content-tag prior and error sparsity," in *Proc. 18th ACM Int. Conf. Multimedia*, 2010, pp. 461–470.
- [2] J. Sang, J. Liu, and C. Xu, "Exploiting user information for image tag refinement," in *Proc. 19th ACM Int. Conf. Multimedia*, 2011, pp. 1129–1132.
- [3] L. Wu, R. Jin, and A. K. Jain, "Tag completion for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 716–727, Mar. 2013.
- [4] J. Tang, X. Shu, G. Qi, Z. Li, M. Wang, S. Yan, and R. Jain, "Tri-clustered tensor completion for social-aware image tag refinement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 8, pp. 1662–1674, Aug. 2017.
- [5] J. Sang, C. Xu, and J. Liu, "User-aware image tag refinement via ternary semantic analysis," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 883–895, Jun. 2012.
- [6] G.-J. Qi, C. C. Aggarwal, Q. Tian, H. Ji, and T. S. Huang, "Exploring context and content links in social media: A latent space method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 850–862, May 2012.
- [7] D. Liu, X.-S. Hua, M. Wang, and H.-J. Zhang, "Image retagging," in *Proc. 18th ACM Int. Conf. Multimedia*, 2010, pp. 491–500.
- [8] Y. Yang, Z. Huang, H. T. Shen, and X. Zhou, "Mining multi-tag association for image tagging," *World Wide Web*, vol. 14, pp. 133–156, 2011.
- [9] X. Xu, L. He, H. Lu, A. Shimada, and R.-I. Taniguchi, "Non-linear matrix completion for social image tagging," *IEEE Access*, vol. 5, pp. 6688–6696, 2017.
- [10] X. Li, B. Shen, B.-D. Liu, and Y.-J. Zhang, "A locality sensitive low-rank model for image tag completion," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 474–483, Mar. 2016.
- [11] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.
- [12] D. Lin, "Using syntactic dependency as local context to resolve word sense ambiguity," in *Proc. 35th Annu. Meeting Assoc. Comput. Linguistics*, 1997, pp. 64–71.
- [13] P. Cui, S.-W. Liu, W.-W. Zhu, H.-B. Luan, T.-S. Chua, and S.-Q. Yang, "Social-sensed image search," *ACM Trans. Intell. Syst. Technol.*, vol. 32, no. 2, 2014, Art. no. 8.
- [14] W. Liu, J. Wang, S. Kumar, and S.-F. Chang, "Hashing with graphs," in *Proc. 28th Int. Conf. Int. Conf. Mach. Learn.*, 2011, pp. 1–8.
- [15] M. Wang, W. Fu, S. Hao, H. Liu, and X. Wu, "Learning on big graph: Label inference and regularization with anchor hierarchy," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 5, pp. 1101–1114, May 2017.
- [16] M. Wang, B. Ni, X.-S. Hua, and T.-S. Chua, "Assistive tagging: A survey of multimedia tagging with human-computer joint exploration," *ACM Comput. Surveys*, vol. 44, no. 4, 2012, Art. no. 25.
- [17] X. Li, T. Uricchio, L. Ballan, M. Bertini, C. G. Snoek, and A. D. Bimbo, "Socializing the semantic gap: A comparative survey on image tag assignment, refinement, and retrieval," *ACM Comput. Surveys*, vol. 49, no. 1, 2016, Art. no. 14.
- [18] J. Fu, J. Wang, Y. Rui, X.-J. Wang, T. Mei, and H. Lu, "Image tag refinement with view-dependent concept representations," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 8, pp. 1409–1422, Aug. 2015.
- [19] H. Xu, J. Wang, X.-S. Hua, and S. Li, "Tag refinement by regularized LDA," in *Proc. 17th ACM Int. Conf. Multimedia*, 2009, pp. 573–576.
- [20] M. Chen, A. Zheng, and K. Weinberger, "Fast image tagging," in *Proc. 30th Int. Conf. Mach. Learn.*, 2013, pp. 1274–1282.
- [21] Z. Li, J. Tang, and T. Mei, "Deep collaborative embedding for social image understanding," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2018, doi: 10.1109/TPAMI.2018.2852750
- [22] X. Liu, S. Yan, T.-S. Chua, and H. Jin, "Image label completion by pursuing contextual decomposability," *ACM Trans. Multimedia Comput., Commun. Appl.*, vol. 8, no. 2, 2012, Art. no. 21.
- [23] Z. Li and J. Tang, "Weakly supervised deep matrix factorization for social image understanding," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 276–288, Jan. 2017.
- [24] Z. Lin, G. Ding, M. Hu, J. Wang, and X. Ye, "Image tag completion via image-specific and tag-specific linear sparse reconstructions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 1618–1625.
- [25] X. Li, Y.-J. Zhang, B. Shen, and B.-D. Liu, "Image tag completion by low-rank factorization with dual reconstruction structure preserved," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 3062–3066.
- [26] D. Rafailidis, A. Axenopoulos, J. Etzold, S. Manolopoulou, and P. Daras, "Content-based tag propagation and tensor factorization for personalized item recommendation based on social tagging," *ACM Trans. Interactive Intell. Syst.*, vol. 3, no. 4, 2014, Art. no. 26.
- [27] Z. Feng, S. Feng, R. Jin, and A. K. Jain, "Image tag completion by noisy matrix recovery," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 424–438.
- [28] C. Deng, R. Ji, W. Liu, D. Tao, and X. Gao, "Visual reranking through weakly supervised multi-graph learning," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2600–2607.
- [29] B. Harwood and T. Drummond, "Fannng: Fast approximate nearest neighbour graphs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 5713–5722.
- [30] I. Suzuki and K. Hara, "Centered KNN graph for semi-supervised learning," in *Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2017, pp. 857–860.
- [31] J. Tang, R. Hong, S. Yan, T.-S. Chua, G.-J. Qi, and R. Jain, "Image annotation by knn-sparse graph-based label propagation over noisily tagged web images," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 2, 2011, Art. no. 14.
- [32] B. Chen, J. Wang, Q. Huang, and T. Mei, "Personalized video recommendation through tripartite graph propagation," in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 1133–1136.
- [33] J. Song, Y. Yang, Z. Huang, H. T. Shen, and J. Luo, "Effective multiple feature hashing for large-scale near-duplicate video retrieval," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1997–2008, Dec. 2013.
- [34] M. Norouzi, A. Punjani, and D. J. Fleet, "Fast exact search in hamming space with multi-index hashing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1107–1119, Jun. 2014.
- [35] W. Liu, C. Mu, S. Kumar, and S.-F. Chang, "Discrete graph hashing," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 3419–3427.
- [36] Q.-Y. Jiang and W.-J. Li, "Scalable graph hashing with feature transformation," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 2248–2254.
- [37] W. Liu, J. He, and S.-F. Chang, "Large graph construction for scalable semi-supervised learning," in *Proc. 27th Int. Conf. Int. Conf. Mach. Learn.*, 2010, pp. 679–686.
- [38] M. Wang, W. Fu, S. Hao, D. Tao, and X. Wu, "Scalable semi-supervised learning by efficient anchor graph regularization," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 7, pp. 1864–1877, Jul. 2016.
- [39] S. Kim and S. Choi, "Multi-view anchor graph hashing," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2013, pp. 3123–3127.
- [40] W. Liu, J. Wang, and S.-F. Chang, "Robust and scalable graph-based semi-supervised learning," *Proc. IEEE*, vol. 100, no. 9, pp. 2624–2638, Sep. 2012.
- [41] B. Xu, J. Bu, C. Chen, C. Wang, D. Cai, and X. He, "EMR: A scalable graph-based ranking model for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 102–114, Jan. 2015.
- [42] Y. Xiong, W. Liu, D. Zhao, and X. Tang, "Face recognition via archetype hull ranking," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 585–592.
- [43] Y. Wu, M. Pei, M. Yang, J. Yuan, and Y. Jia, "Robust discriminative tracking via landmark-based label propagation," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1510–1523, May 2015.
- [44] I. S. Dhillon, "Co-clustering documents and words using bipartite spectral graph partitioning," in *Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2001, pp. 269–274.
- [45] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [46] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
- [47] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, 1999, Art. no. 788.
- [48] J. Tang, X. Shu, Z. Li, G.-J. Qi, and J. Wang, "Generalized deep transfer networks for knowledge propagation in heterogeneous domains," *ACM Trans. Multimedia Comput., Commun. Appl.*, vol. 12, no. 4s, 2016, Art. no. 68.
- [49] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, "NUS-WIDE: A real-world web image database from National University of Singapore," in *Proc. ACM Int. Conf. Image Video Retrieval*, 2009, Art. no. 48.