Computer Vision and Image Understanding 124 (2014) 71–78

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Projective Matrix Factorization with unified embedding for social image tagging

Zechao Li [a,b], Jing Liu [b,*], Jinhui Tang [a], Hanqing Lu [b]

[a] School of Computer Science, Nanjing University of Science and Technology, Nanjing, China
[b] National Laboratory of Pattern Recognition, Institute of Automation Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Article history:
Received 9 August 2013
Accepted 1 February 2014

Keywords:
Projective Matrix Factorization
Unified embedding
Structure preserving
Social tagging

ABSTRACT

This paper presents a general formulation, named ProJective Matrix Factorization with unified embedding (PJMF), by which social image retagging is transformed to the nearest tag-neighbor search for each image. We solve the proposed PJMF as an optimization problem mainly considering the following issues. First, we attempt to find two latent representations in a unified space for images and tags respectively and explore the two representations to reconstruct the observed image-tag correlation in a nonlinear manner. In this case, the relevance between an image and a tag can be directly modeled as the pair-wise similarity in the unified space. Second, the image latent representation is assumed to be projected from its original visual feature representation with an orthogonal transformation matrix. The projection makes convenient to embed any images including out-of-samples into the unified space, and naturally the image retagging problem can be solved by the nearest tag-neighbors search for those images in the unified space. Third, local geometry preservations of image space and tag space respectively are explored as constraints in order to make image similarity (and tag relevance) consistent in the original space and the corresponding latent space. Experimental results on two publicly available benchmarks validate the encouraging performance of our work over the state-of-the-arts.

1. Introduction

With the permeation of Web 2.0, there are explosive photo sharing websites with large-scale image collections available online, such as Flickr and Picasa. The Web 2.0 websites allow users not only share their photos, but tag and comment their interested ones. Due to the subjectivity and diversity of such social tagging behaviors, noisy and missing tags for images are inevitable, which limits the performance of tag-based image retrieval system. Thus, it is necessary and challenging to retag the large-scale images precisely by leveraging those possibly imprecise tagging information.

Image annotation as a special case of image retagging is a classical task of computer vision and pattern recognition and numerous studies have been exploited [1–6]. Despite being studied extensively, most of regular image annotation approaches fail to handle large-scale social image tagging tasks since they are usually designed on small-scale manually-labeled data. Besides, due to the diversity of knowledge and cultural background of users, social tagging is often subjective and inaccurate. Moreover, existing studies reveal that many tags provided by Flickr users are imprecise and there are only around 50% tags actually related to the image [7,8]. Consequently, the tags associated with social images could be noisy, irrelevant and incomplete, which may severely deteriorate the performance of text-based image retrieval [8]. Some previous image retagging methods [9,8,10,11] have been proposed to refine and complement the tagging information of social images in a transductive learning manner. That is, most of them cannot directly handle the new images, i.e., the out-of-sample problem. Therefore, how to retag large-scale social images with noisy or missing tags while make the learned model competent for the new image tagging problem, is an urgent task.

To this end, in this paper, we propose a novel matrix factorization approach, named ProJective Matrix Factorization with unified embedding (PJMF) for tag relevance learning, and apply it to social image retagging task. The framework of our solution is illustrated in Fig. 1. During the learning process, the original tagging information of social images is used to calculate the image-tag correlation matrix. The two low-dimensional latent representations in a unified space for images and tags respectively are learned to reconstruct the observed correlation matrix with minimum errors in an optimization problem. Different from the traditional matrix factorization approaches, the embedding makes image points and tag points in the unified space comparable, and

* Corresponding author.
E-mail address: jliu@nlpr.ia.ac.cn (J. Liu).

http://dx.doi.org/10.1016/j.cviu.2014.02.001
1077-3142/

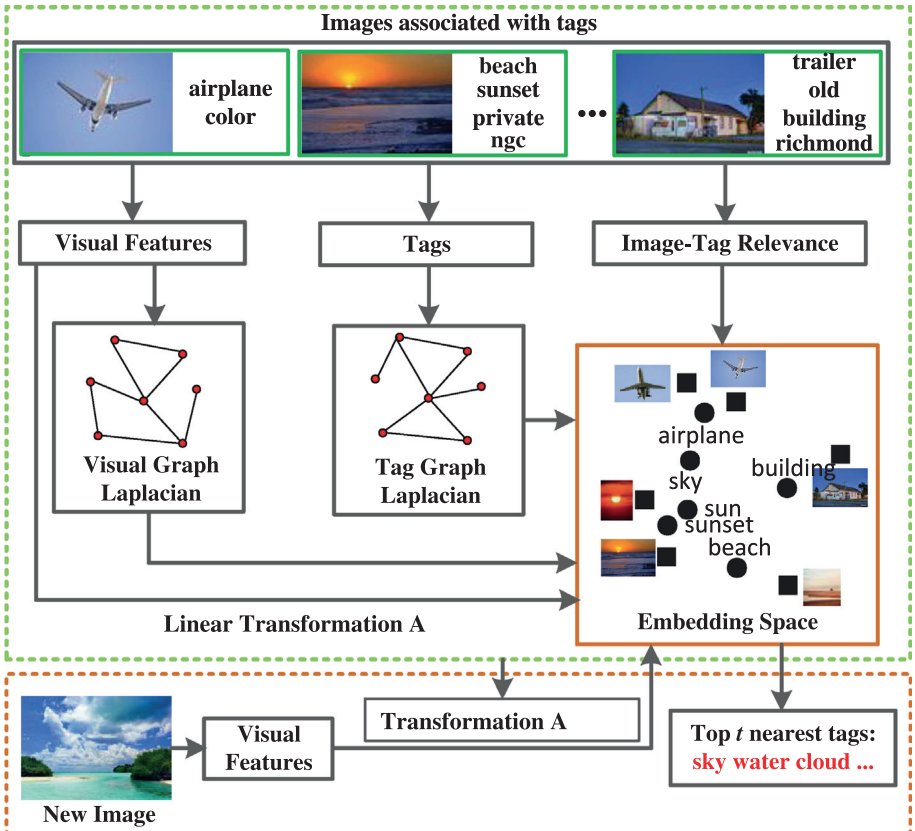

Fig. 1. The proposed framework of our work. As a result of learning, the image latent representation projected with a transformation matrix **A** and the tag latent representation are embedded in a unified space. Given an image even an untagged one, we project it into the embedding space via **A** and tag the image by the top t nearest tags.

naturally transforms the image tagging problem to the nearest tag-neighbors search. We also formulate another two issues into the optimization problem for social image tagging task. We assume that the latent image representation is an explicit projection from original image representation to the unified space via an orthogonal transformation matrix. This provides a solution to the out-of-sample representations for images. Besides, to preserve the local geometric information for image space and tag space, two constraints are imposed on the corresponding factor matrices. Obtaining the learned transformation matrix and the tag latent representation, we can easily annotate any image even an untagged image by first embedding it into the unified space via the transformation matrix and then searching its neighbors among the tag latent representations. Finally, we conduct extensive experiments on two publicly available benchmarks, and demonstrate the superiority of the proposed PJMF over the state-of-the-art methods.

In summary, our main contributions are given as follows.

- A general formulation for matrix factorization is proposed, by which the two latent representations for images and tags are embedded into a unified space. This makes the latent points comparable and transforms the image tagging problem by nearest tag-neighbors search.
- The projection from the original image feature space to the unified embedding space is incorporated into the matrix factorization framework, which provides an entrance to the out-of-sample problem.
- The local geometrical properties of image space and tag space respectively are preserved for discovering the latent embedding space.

The remaining content is organized as follows. Section 2 overviews related work. We present the proposed matrix factorization approach in Section 4. Section 5 discusses data sets and experimental settings. In Section 6, experimental results and analysis are reported. Conclusions are discussed in Section 7.

2. Previous work

2.1. Matrix factorization

The goal of Matrix Factorization (MF) is to decompose the data as the product of two low-dimensional factor matrices. Various proposals about MF differ in the constraints that are imposed on the factorization, and the measures of approximate error.

The most common form of MF is finding a low-rank approximation to a fully observed data by minimizing the sum-squared difference to it. The two best-known approaches are PCA and SVD, which restrict the factor matrices to be orthogonal. Another popular MF is Non-negative MF (NMF) [12], which requires the data and the factor matrices to be non-negative. There are a number of variants of NMF methods [13–16]. In [17], Projective NMF (PNMF) was introduced which approximates a data matrix by its nonnegative subspace projection. Recently, Probabilistic Matrix Factorization (PMF) [18] is proposed to handle the sparse and imbalanced datasets, which is based on the Gaussian (normal) assumption for both the prior and the likelihood term. Multi-correlation PMF (MPMF) [4] and Correlation Consistent PMF (CCPMF) [10] are proposed to jointly exploit multiple correlations simultaneously. Maximum Margin MF (MMMF) [19] is another MF approach, which adopts low-norm instead of low-rank factorization.