Neurocomputing 172 (2016) 19–26

Contents lists available at ScienceDirect

Neurocomputing

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

Projective nonnegative matrix factorization for social image retrieval

Qiuli Liu [a], Zechao Li [b,*]

[a] Capital University of Economics and Business, No. 121 Zhangjialukou Road, Fengtai District, Beijing 100070, PR China
[b] School of Computer Science, Nanjing University of Science and Technology, No. 200, Xiaolinwei Road, Nanjing 210094, PR China

ARTICLE INFO

Article history:
Received 30 October 2013
Received in revised form
18 September 2014
Accepted 23 September 2014
Available online 14 May 2015

Keywords:
Nonnegative matrix factorization
Latent subspace
Image retrieval

ABSTRACT

Increasingly many social images with tags are available on photo sharing websites. Due to the subjectivity and diversity of social tagging behaviors, noisy and missing tags for images are inevitable. To tackle this problem, this paper proposes a novel factor analysis model, named ProjecTive Nonnegative Matrix Factorization (PTNMF) with $\ell_{2,1}$-norm regularization, which introduces linear transformation and $\ell_{2,1}$-norm minimization into a joint framework of NMF. For tagging data, a new interpretation is adopted to distinguish the relevant tags and irrelevant tags instead of the typically used binary scheme. In our model, the image latent representation is assumed to be projected from its original feature representation with an orthogonal transformation matrix. The projection makes convenient to embed any images including out-of-samples into the latent space. That is, the proposed method enables to handle the out-of-sample problem. The $\ell_{2,1}$-norm regularization makes the transformation matrix suitable for selecting the effective features. Local geometry preservations of image space (tag space) are explored as constraints in order to make image similarity (tag correlation) consistent in the original space and the corresponding latent space. We investigate the performance of the proposed method on image retrieval and compare it to existing work on the challenging NUS-WIDE dataset. Extensive experiments indicate the effectiveness and potentials of the proposed method in real-world applications.

1. Introduction

With the permeation of Web 2.0 technologies and digital cameras, there is an explosion of social media sharing system available online, such as Flickr, Facebook and Zooomr. Users can share their photos, tag and comment others' ones. Tagging, in general, allows users to describe an image with a list of tags, which can be utilized to search, browse and organize images. However, due to the subjectivity and diversity of amateur tagging, tags are known to be ambiguous, limited in terms of completeness, and overly personalized [1,2]. As a consequence, user-provided tags are often imprecise, biased and incomplete for describing the content of the images. Thus, during recent years, it has attracted much research attention to reliably learn the relevance of a tag with respect to the visual content it is describing [3,4], which is an essential issue for image retrieval.

Many efforts have been made to learn the relevance of tags to images. One category of tag relevance learning methods is to predict relevant tags for images with no tag [3–11]. One of the related works is Multi-correlation Probabilistic Matrix Factorization (MPMF) model [9], which is to combine the inter- and intra-correlation matrices by the shared latent matrices based on Probabilistic Matrix Factorization (PMF) [12]. They are mostly train on manually labeled data and tested on small data sets [13], which make them unsuitable for social image tagging. In the second scenario, given an image labeled with some tags, tag relevance learning can be used to remove noisy tags, recommend new relevant tags or reduce tag ambiguity [2,14–19]. Tag ranking [15] exploits pairwise similarity between tags by random walk to refine the ranking score. In [16,17], both the image similarity and tag correlation are exploited simultaneously to discover the tag relevance. In [16], the image tags are refined by adding consistent constraints on Robust Principal Component Analysis (RPCA) [20]. However, they cannot handle the out-of-sample problem and do not utilize the visual feature directly. That is, massive new images can not be tagged.

To this end, in this paper, we tackle the social image tag relevance task by the Nonnegative Matrix Factorization (NMF) model. Different from the above MF-based methods, we propose a novel NMF algorithm, namely ProjecTive Nonnegative Matrix Factorization (PTNMF) with $\ell_{2,1}$-norm regularization, to collaboratively predict the tag relevance. To handle missing tags and noisy tags, a new interpretation scheme is proposed motivated by [21]. To handle the out-of-sample problem, we assume that the latent image representation is an explicit projection from original image representation via an orthogonal

* Corresponding author.
E-mail address: zechao.li@gmail.com (Z. Li).

http://dx.doi.org/10.1016/j.neucom.2014.09.094
0925-2312/

transformation matrix. The $\ell_{2,1}$-norm regularization is introduced to learn a reliable transformation matrix [22]. To preserve the local geometrical properties in both visual space and semantic space, the visual similarity and tag relevance are explored jointly and simultaneously. We also propose an efficient iterative algorithm to optimize the problem. Experiments on different social image datasets demonstrate that our algorithm outperforms the state-of-the-art algorithms.

The main contributions of this paper are summarized as follows:

1. We propose a novel ProjecTive Nonnegative Matrix Factorization (PTNMF) algorithm, which can handle the out-of-sample problem with the assumption that the image latent representation is projected from its original feature representation through an orthogonal transformation matrix.
2. To handle the irrelevant visual features, an $\ell_{2,1}$-norm regularization is introduced to learn a reliable transformation matrix which is suitable for selection of the effective features.
3. To keep the local geometrical properties in both visual and semantic spaces, the visual similarity and tag relevance are explored jointly and simultaneously.

The reminder of this paper is organized as follows. We review related work in Section 2. Section 3 elaborates the proposed nonlinear matrix factorization with unified embedding algorithm. In Section 4, extensive experiments are conducted to evaluate the performance of the proposed method and compare it to other related methods. The conclusion of this paper with future work discussion is presented in Section 5.

2. Related work

It is an essential issue to estimate the relevance of tags with respect to images in text-based image retrieval. The related techniques are categorized into two main scenarios, namely tag annotation for untagged images and tag refinement for tagged images.

Methods in the first category predict relevant tags for images with no tag. A variety of methods have been proposed to annotate images automatically [5,7,9,23–30], which can be categorized into two main types, *generative models* and *classification models*. The generative models try to estimate the probabilistic relationship between tags and images. By assigning relevant scores of tags to images, the annotated results can be utilized to help the task of image retrieval.

In the second scenario, given an image labeled with some tags, tag relevance learning can be used to remove noisy tags, recommend new relevant tags or reduce tag ambiguity. Many approaches have been proposed to tackle the tag relevance learning problem [2,14–17,31–39]. The Random Walk with Restarts (RWR) algorithm [31] is proposed to leverage the co-occurrence-based tag similarity and the information of the original annotated order of tags. The tag refinement problem is formulated as a Markov process and the candidate tags are treated as the states in [14]. In [2], a neighbor voting algorithm is proposed to estimate a tag's relevance by exploiting tagging redundancies among multiple users. The tag relevance is determined based on the number of such votes from the nearest neighbors. Tag ranking [15] further exploits pairwise similarity between tags by random walk to refine the ranking score. In [9,16,17], both the image similarity and tag correlation are exploited simultaneously to discover the tag relevance. A Multi-correlation Probabilistic Matrix Factorization (MPMF) model [9] is proposed to combine the inter- and intra-correlation matrices by the shared latent matrices. In [16], the image labels are refined by decomposing the observed label matrix into a low-rank refined matrix and a sparse error matrix. A two-view learning approach is proposed to address the tag ranking problem in [17]. A shared subspace learning framework based on NMF is proposed to leverage a secondary source to improve retrieval performance from a primary dataset in [40]. In [39], a unified subspace is learned for images and tags, which can refine tags of images by nearest tags search in the underlying subspace. However, the above methods focus on refining images tags. They cannot assign tags to the new images. That is, they cannot address the out-of-sample problem.

Different from the previous work, this paper presents a novel projective nonnegative matrix factorization approach to estimate the relevance of tags to social images. The local visual geometry in image space and local textual geometry in tag space are exploited simultaneously. This method can be employed to tag and refine images.

3. The proposed PTNMF algorithm

3.1. Nonnegative matrix factorization

Before getting started, we first summarize some notations. Throughout this paper, we use bold uppercase characters to denote matrices, bold lowercase characters to denote vectors. For an arbitrary matrix $\mathbf{M}$, $\mathbf{m}_i$ means the ith column vector of $\mathbf{M}$, M_{ij} denotes the (i,j)th entry of $\mathbf{M}$ and $\mathrm{Tr}[\mathbf{M}]$ is the trace of $\mathbf{M}$ if $\mathbf{M}$ is square. $\mathbf{M}^T$ is the matrix transposition operation. For any $\mathbf{M} \in \mathcal{R}^{r \times t}$, its $\ell_{2,1}$-norm is defined as

$$\|\mathbf{M}\|_{2,1} = \sum_{i=1}^{r} \sqrt{\sum_{j=1}^{t} M_{ij}^2} = \mathrm{Tr}[\mathbf{M}^T\mathbf{D}\mathbf{M}]. \tag{1}$$

Here $\mathbf{D}$ is a diagonal matrix with $D_{ii} = \frac{1}{\|\mathbf{m}^i\|_2}$.

Assume that we have n tagged images and m tags. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$ denote the data sample set, in which $\mathbf{x}_i \in \mathcal{R}^d$ denotes the feature descriptor of the ith sample. $\mathbf{Y} \in \mathcal{R}^{m \times n}$ denotes the tag-image associated matrix, such that $Y_{ij} = 1$ if $\mathbf{x}_i$ is tagged by the jth tag, and 0 otherwise. The task of NMF is to derive two nonnegative factor matrices $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m] \in \mathcal{R}_+^{p \times m}$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n] \in \mathcal{R}_+^{p \times n}$, where $p < \min(m, n)$. That is, given $\mathbf{Y} \in \mathcal{R}^{m \times n}$, a low-rank NMF approach seeks to approximate it by a multiplication of two p-rank nonnegative factors

$$\min_{\mathbf{U},\mathbf{V}} \|\mathbf{Y} - \mathbf{U}^T\mathbf{V}\|_F^2 \quad \text{s.t.} \quad \mathbf{U}, \mathbf{V} \geq 0, \tag{2}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. To optimize the objective, an iterative multiplicative updating algorithm was proposed in [41] as follows:

$$V_{mj} \leftarrow V_{mj} \frac{(\mathbf{U}\mathbf{Y})_{mj}}{(\mathbf{U}\mathbf{U}^T\mathbf{V})_{mj}} \tag{3}$$

$$U_{mi} \leftarrow U_{mi} \frac{(\mathbf{V}\mathbf{Y}^T)_{mi}}{(\mathbf{V}\mathbf{V}^T\mathbf{U})_{mi}} \tag{4}$$

3.2. The objective function

Since the original tagging data contains noisy tags and missing tags, it is unreasonable and incorrect to treat the tagged tags equally and encode all the unobserved data as 0 in the binary interpretation [9,16,17]. To address the above problems, a novel interpretation scheme is presented based on the context and semantic information. It is reasonable to assume that the tags that are highly relevant in terms of the context and semantic information are likely to appear in the same images. On the other hand, one tag that is irrelevant to other tags in the same image may well