Pattern Recognition Letters 98 (2017) 68–74

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Adaptive total-variation for non-negative matrix factorization on manifold

Chengcai Leng[a,b,c], Guorong Cai[c,d,*], Dongdong Yu[c], Zongyue Wang[d]

[a] School of Mathematics, Northwest University, Xi'an 710127, China
[b] School of Mathematics and Information Sciences, Nanchang Hangkong University, Nanchang 330063, China
[c] Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[d] College of Computer Engineering, Jimei University, Xiamen 361021, China

ARTICLE INFO

Article history:
Received 20 September 2016
Available online 26 August 2017

JEL classification:
41A05
41A10
65D05
65D17

Keywords:
Adaptive total variation
Non-negative matrix factorization
Manifold learning

ABSTRACT

Non-negative matrix factorization (NMF) has been widely applied in information retrieval and computer vision. However, its performance has been restricted due to its limited tolerance to data noise, as well as its inflexibility in setting regularization parameters. In this paper, we propose a novel sparse matrix factorization method for data representation to solve these problems, termed Adaptive Total-Variation Constrained based Non-Negative Matrix Factorization on Manifold (ATV-NMF). The proposed ATV can adaptively choose the anisotropic smoothing scheme based on the gradient information of data to denoise or preserve feature details by incorporating adaptive total variation into the factorization process. Notably, the manifold graph regularization is also incorporated into NMF, which can discover intrinsic geometrical structure of data to enhance the discriminability. Experimental results demonstrate that the proposed method is very effective for data clustering in comparison to the state-of-the-art algorithms on several standard benchmarks.

1. Introduction

Matrix Factorization (MF) plays the fundamental role in various emerging applications ranging from information retrieval to data mining [1]. It typically adopts a sparse representation to obtain low-dimensional matrix, which can deal with many classical classification and clustering problems efficiently and robustly [2–4]. In order to avoid the curse of dimensionality, different forms of dimensionality reduction schemes like Principal Component Analysis (PCA), ISOMAP [5], Locally Linear Embedding (LLE) [6], Laplacian Eigenmap [7] and Isometric Projection [8]. NMF [9] incorporates the non-negativity constraint to achieve a parts-based representation.

NMF allows only additive, not subtractive, combination of the original data, and which is effective to capture the underlying structure of the data combining non-negative constraints in a parts-based low dimensional space. Usually, the rank of the NMF is generally chosen so that the matrix factorization can be regarded as a compressed form of the data [9,10]. NMF has been widely used for clustering [11,12], face recognition [13–15] and image or data analysis [2,16]. To overcome the difficulty in modeling the intrinsic geometrical structure, Manifold learning [4–6,17,18] has been introduced into NMF. For instance, Cai et al. [19] presented a graph regularized NMF (GNMF) by adding a graph manifold term to NMF, While promising, manifold-based NMF is typically sensitive to data noise.

Since NMF model does not consider noise signal, its performance has been restricted due to the fact that it is very hard to determine appropriate regularization parameters. In order to resolve these problems, we will add an adaptive total variation regularization item to NMF model. It is worth noting that Total variation (TV), first introduced by Rudin et al. [20], is effective for piecewise constant reconstruction, thus can preserve the boundary of large objects well. Since then TV regularization has been widely used for denoising tasks in image processing, computer vision and image reconstruction, such as data representation [21], face recognition [15,22]. To this end, total variation scheme has been proposed to handle data noise by combining TV term [23,24]. However, TV based NMF cannot well discover and reveal the intrinsic geometrical and structure information of data and it is difficult to fix the TV regularization parameter of TV term.

In this paper, we present a novel NMF scheme that correctly handles the data noise as well as modeling the intrinsic geometric structure of data, terms Adaptive Total-Variation Constrained based Non-negative Matrix Factorization on Manifold (ATV-NMF). First, in order to discover intrinsic geometrical structure, we

* Corresponding author at: College of Computer Engineering, Jimei University, Xiamen 361021, China.
E-mail address: guorongcai.jmu@gmail.com (G. Cai).

http://dx.doi.org/10.1016/j.patrec.2017.08.027
0167-8655/

incorporate the graph regularization to NMF. Second, the Adaptive Total-Variation (ATV) regularization is incorporated to choose adaptively the anisotropic smoothing scheme based on the data gradient to denoise or preserve adaptively the feature details. ATV also avoids choosing the regularization parameter to enhance the discrimination ability. Finally, we present a novel iterative update rule that achieve ATV-NMF. Experimental results show that the proposed method is better compared to the state-of-the-art schemes for data clustering.

The rest of this paper is organized as follows: In Section 2, we propose the ATV-NMF on manifold method. Section 3 presents experimental results and Section 4 gives conclusions and future work.

2. ATV-NMF On manifold

In this section, we first describe the basic idea of ATV method. In principle, ATV and graph regularization is introduced into NMF to preserve edge or details, as well as to discover and enhance the intrinsic geometrical data structure to improve the discriminability. As for data clustering, the database is regarded as an $m \times n$ matrix V, each column of which contains m non-negative values of one of the n images. Then the task of ATV-NMF is to construct approximate factorizations of the form $V = WH$, where W and H are respectively $m \times r$ and $r \times n$ matrix factors, and r denotes the rank of the factorization.

2.1. Adaptive total-variation

Our ATV-NMF model is inspired by the adaptive total variation regularization proposed in [25] so that the proposed model can adaptively choose the anisotropic smoothing scheme based on the gradient information of data to denoise or preserve feature details, which can be defined as:

$$E(H) = ||H||_{ATV} \tag{1}$$

where E is the energy function of H, $||H||_{ATV} = \int_\Omega \frac{1}{p(x,y)} |\nabla H|^{p(x,y)} dxdy$ denotes the adaptive TV regularization term, $p(x, y) = 1 + \frac{1}{1+|\nabla H|^2}$, $1 < p(x,\ y) < 2$, $(\nabla H)(i, j) = ((\partial_x H)(i, j), (\partial_y H)(i, j))$ is a discrete gradient form with $(\partial_x H)(i, j)$ and $(\partial_y H)(i, j)$, given as follows:

$$(\partial_x H)(i, j) = \begin{cases} H(i+1, j) - H(i, j) & if\ i < r \\ H(1, j) - H(r, j) & if\ i = r \end{cases}$$

$$(\partial_y H)(i, j) = \begin{cases} H(i, j+1) - H(i, j) & if\ j < n \\ H(i, 1) - H(i, n) & if\ j = n \end{cases}$$

The adaptive TV regularization including a diffusion coefficient $\frac{1}{|\nabla H|^{2-p}}$ in Eq. (8), which is used to control the speed of the diffusion based on the gradient information. For edges, $|\nabla H|^{2-p}$ has big values, the $\frac{1}{|\nabla H|^{2-p}}$ is small and the diffusion is very weak along the edge directions, which helps preserve edges. In a smooth region, $|\nabla H|^{2-p}$ has small values, the $\frac{1}{|\nabla H|^{2-p}}$ is big and the diffusion is strong, which helps in denoising. In addition, the ATV model has some fundamental properties, which has numerical stability solution, can avoid the staircase effect, and is able to preserve or enhance finer scale data features, such as edges or textures, while denoising [25].

2.2. Multiplicative updating rules

Using the ATV as the regularization term, the refined ATV-NMF model is designed by solving the following objective function:

$$\begin{aligned} O_{ATV-NMF} = {} & ||V - WH||_F^2 + \lambda Tr(HLH^T) \\ & +2||H||_{ATV}. \qquad s.t.\ W \geq 0, H \geq 0 \end{aligned} \tag{2}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, $\lambda \geq 0$ is a regularization parameter, $Tr(.)$ denotes the trace of a matrix, S is the weight matrix whose entry S_{ij} measures the similarity between each vertex pair $(v_i,\ v_j)$, D is a diagonal matrix with column sums of S as its diagonal entries. i.e., $D_{ij} = \sum_{i=1}^n S_{ij}$, $L = D - S$ is called graph Laplacian matrix [26].

Since the objective function $O_{ATV-NMF}$ in Eq. (2) is not convex in W and H, we therefore resort to an iterative updating algorithm to obtain an approximate optimal solution of $O_{ATV-NMF}$. In order to obtain the solution of the objective function $O_{ATV-NMF}$ in Eq. (2), we need to find an iterative updating algorithm to achieve the minimization of $O_{ATV-NMF}$ by gradient descent algorithm [27]. The gradient of the objective function $O_{ATV-NMF}$ with respect to W and H are given as follows:

$$\frac{\partial O_{ATV-NMF}}{\partial W_{i,l}} = -2(VH^T - WHH^T)_{i,l} \tag{3}$$

$$\frac{\partial O_{ATV-NMF}}{\partial H_{l,j}} = -2\left(W^TV - W^TWH - \lambda HL + div\left(\frac{\nabla H}{|\nabla H|^{2-p}}\right)\right)_{l,j} \tag{4}$$

The additive update rules for problem (2) by Eqs. (3) and (4) can be obtained as follows:

$$W_{i,l} \leftarrow W_{i,l} + \xi_{i,l}(VH^T - WHH^T)_{i,l} \tag{5}$$

$$H_{l,j} \leftarrow H_{l,j} + \eta_{l,j}\left(W^TV - W^TWH - \lambda HL + div\left(\frac{\nabla H}{|\nabla H|^{2-p}}\right)\right)_{l,j} \tag{6}$$

where $\xi_{i,l} = \frac{W_{i,l}}{(WHH^T)_{i,l}}$ and $\eta_{l,j} = \frac{H_{l,j}}{(W^TWH + \lambda HD)_{l,j}}$ are the step sizes of the updates, and the multiplicative updating rules can be formulated as follows:

$$W_{i,l} \leftarrow W_{i,l} \frac{\left(VH^T\right)_{i,l}}{\left(WHH^T\right)_{i,l}} \tag{7}$$

$$H_{l,j} \leftarrow H_{l,j} \frac{(W^TV + \lambda HS + div(\frac{\nabla H}{|\nabla H|^{2-p}}))_{l,j}}{(W^TWH + \lambda HD)_{l,j}} \tag{8}$$

where div denotes the divergence, i.e., $div = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$, $\nabla H = (\partial_x H, \partial_y H)$ denotes the gradient, and $|\nabla H| = \sqrt{(\partial_x H)^2 + (\partial_y H)^2}$ is the norm of the gradient. The similar form of the Eq. (8) can be found in [22], and the discrete form of the $div(\frac{\nabla H}{|\nabla H|^{2-p}})$ can also be found based on the operator of the divergence and the gradient by using total variation principal [25]. The derivation of Eq. (8) is given as belows.

Note that Eq. (6) is the additive update rule, where $\eta_{l,j} = \frac{H_{l,j}}{(W^TWH + \lambda HD)_{l,j}}$. Let $L = D - S$ be the graph Laplacian matrix [26], thus we have:

$$H_{l,j} \leftarrow H_{l,j} + \eta_{l,j}\left(W^TV - W^TWH - \lambda HL + div\left(\frac{\nabla H}{|\nabla H|^{2-p}}\right)\right)_{l,j}$$

$$H_{l,j} \leftarrow H_{l,j} + \eta_{l,j}\left(W^TV - W^TWH - \lambda H(D - S) + div\left(\frac{\nabla H}{|\nabla H|^{2-p}}\right)\right)_{l,j}$$

$$H_{l,j} \leftarrow H_{l,j} + \eta_{l,j}\left(-W^TWH - \lambda HD + W^TV + \lambda HS + div\left(\frac{\nabla H}{|\nabla H|^{2-p}}\right)\right)_{l,j}$$

$$H_{l,j} \leftarrow H_{l,j} + \eta_{l,j}(-W^TWH - \lambda HD)_{l,j} + \eta_{l,j}\left(W^TV + \lambda HS + div\left(\frac{\nabla H}{|\nabla H|^{2-p}}\right)\right)_{l,j}$$

$$H_{l,j} \leftarrow \eta_{l,j}\left(W^TV + \lambda HS + div\left(\frac{\nabla H}{|\nabla H|^{2-p}}\right)\right)_{l,j}$$