Pattern Recognition 44 (2011) 2800–2810

Contents lists available at ScienceDirect

Pattern Recognition

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

Nonlinear nonnegative matrix factorization based on Mercer kernel construction

Binbin Pan [a], Jianhuang Lai [b,*], Wen-Sheng Chen [c]

[a] School of Mathematics and Computational Science, Sun Yat-Sen University, Guangzhou, China
[b] School of Information Science and Technology, Sun Yat-Sen University, Guangzhou, China
[c] College of Mathematics and Computational Science, Shenzhen University, Shenzhen, China

ARTICLE INFO

Article history:
Received 28 May 2010
Received in revised form
10 March 2011
Accepted 16 March 2011
Available online 6 April 2011

Keywords:
Nonnegative matrix factorization
Mercer kernel
Kernel mapping
Face recognition

ABSTRACT

Generalizations ofnonnegative matrix factorization (NMF) in kernel feature space, such as projected gradient kernel NMF (PGKNMF) and polynomial Kernel NMF (PNMF), have been developed for face and facial expression recognition recently. However, these existing kernel NMF approaches cannot guarantee the nonnegativity of bases in kernel feature space and thus are essentially semi-NMF methods. In this paper, we show that nonlinear semi-NMF cannot extract the localized components which offer important information in object recognition. Therefore, nonlinear NMF rather than semi-NMF is needed to be developed for extracting localized component as well as learning the nonlinear structure. In order to address the nonlinear problem of NMF and the semi-nonnegative problem of the existing kernel NMF methods, we develop the nonlinear NMF based on a self-constructed Mercer kernel which preserves the nonnegative constraints on both bases and coefficients in kernel feature space. Experimental results in face and expressing recognition show that the proposed approach outperforms the existing state-of-the-art kernel methods, such as KPCA, GDA, PNMF and PGKNMF.

1. Introduction

Nonnegative matrix factorization (NMF) [1,2], which aims to find part-based representation of nonnegative data, is an unsupervised subspace method. It decomposes the data into two nonnegative matrices,[1] the bases and the coefficients, in which the data are represented as a non-subtractive combination of bases. The nonnegativity constraints are compatible with the intuitive notion of combining parts to form a whole. For example, the bases represent the parts of face (nose, eyes, etc.) in face representation, and these parts are combined together to compose the face. Therefore, NMF is a promising approach for the extraction of localized components. Due to intuitive interpretability and part-based representation, NMF and its alternatives have been widely applied to face recognition [3], multimedia signal processing [4], document clustering [5], environmetrics [6], chemometrics [7], and bioinformatics [8].

Classical NMF is a linear model and it may fail to discover the nonlinearities of data. However, many real-life data have latent nonlinear structures. For example, the distribution of face image variations under different pose and illumination is complex and nonlinear. Therefore, the performance of traditional NMF is limited. Accordingly, it is necessary to develop the nonlinear NMF. We suggest here using the combination of kernel technology and NMF for this purpose. Kernel method is a powerful technique in handling nonlinear correlations, and it has been widely used for the extension of linear method to nonlinear version. The idea of kernel method is to map the data nonlinearly into a kernel feature space, where the nonlinearities will be linearized. Then, the linear method can be performed in the kernel feature space to process the nonlinear data. The great success of kernel to model the nonlinearities attracts many researchers for in-depth exploring [9]. For example, kernel principal component analysis (KPCA) [10] and generalized discriminant analysis (GDA) [11] were found to outperform their linear versions in different applications.

Recently, generalizations of NMF in kernel feature space, namely polynomial kernel NMF (PNMF) [12] and projected gradient kernel NMF (PGKNMF) [13], have been introduced to model NMF nonlinearly. They learn more useful latent features. Similar to NMF, PNMF approximates the embedded data as the linear combination of bases in kernel feature space by minimizing the squared Euclidean distance. It develops multiplicative updating rules that guaranteed the non-increasing evolution of the cost function. But the updating algorithm is not guaranteed to converge to the stationary points. Moreover, only the polynomial kernels are usable in PNMF. Using projected gradient method, PGKNMF successively optimizes two subproblems [14], which ensures that

* Corresponding author. Tel./fax: +86 20 84110175.
E-mail addresses: alt26cn@gmail.com (B. Pan), stsljh@mail.sysu.edu.cn, sunnyweishi@gmail.com (J. Lai), chenws@szu.edu.cn (W.-S. Chen).

[1] A nonnegative matrix means each entry of the matrix is nonnegative.

0031-3203/$ - see front matter
doi:10.1016/j.patcog.2011.03.023

the limit point is a stationary point and that arbitrary positive kernels can be used. However, neither PNMF nor PGKNMF guaranteed the nonnegative constraints on bases in kernel feature space (cf. Section 3.1). Therefore, PNMF and PGKNMF are essentially semi-nonnegative matrix factorization (semi-NMF) [15][2] other than NMF. As shown in [15], semi-NMF performs worse than NMF in terms of clustering accuracies. Furthermore, to the best of our knowledge, no literature has studied the part-based representation or the classification power of semi-NMF.

Thus, we pose the following two questions:

1. Is semi-NMF fit for object recognition?
2. If not, how to develop nonlinear NMF?

The above two questions will be answered in this paper in the application of face recognition. For the first question, we theoretically illustrate that the bases of semi-NMF do not exhibit the characteristics of sparsity (cf. Section 3.2) which is important for localized component extraction and object recognition [3,16,17]. We also empirically compare the subspace representation and the classification power of semi-NMF with those of NMF. The empirical findings are in accord with the theoretical analysis.

For the second question, different from the previous kernelized methods, this paper develops a novel kernel mapping to impose the nonnegativity. The kernel function induced by the mapping is proven to be a Mercer one. The effectiveness of the proposed algorithm is evaluated by face and facial expression recognition. The results are encouraging, manifesting that nonlinear NMF has superiority over nonlinear semi-NMF.

The rest of this paper is organized as follows. Section 2 introduces the related work briefly. Section 3 presents the shortcomings of PNMF and PGKNMF, which motivates the work of this paper. The details of the proposed algorithm are described in Section 4. Experimental comparisons are given in Section 5, and the conclusions are drawn in Section 6.

2. Related work

In this section, we briefly describe the original NMF method [1] and the related extending works [12,13].

2.1. Nonnegative matrix factorization (NMF)

Give a set of facial images $\{I_i\}_{i=1}^n$, where n is the number of images. By stacking pixels of each image into a column vector x_i, one can get a training set $X=\{x_1,\ldots,x_n\}\subset\Gamma$, where $x_i\in\mathbb{R}^m$, m is the number of pixels for an image, Γ denotes the input space. The x_i's are concatenated to form a matrix $V=[x_1,\ldots,x_n]\in\mathbb{R}^{m\times n}$. NMF [1] aims to find an approximate decomposition

$$V\approx WH \tag{1}$$

by imposing the nonnegative constraints on W and H, where $W\in\mathbb{R}^{m\times r}$ is the bases, $H\in\mathbb{R}^{r\times n}$ is the coefficients, r is the number of bases. These constraints offer some degree of sparsity in bases and coefficients which will be illustrated in Section 3.2. The (1) can be casted as the problem of minimizing the reconstruction error under the nonnegative constraints:

$$E_{NMF}(W,H)=\|V-WH\|_F^2=\sum_{ij}(V_{ij}-(WH)_{ij})^2 \quad \text{s.t. } W,H\geq 0, \tag{2}$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

It is unrealistic to find the global minima of problem (2) since this problem is not convex for variables W and H. Lee and Seung [2] employed coordinate-descent method and ensured that the objective function was non-increasing after each iteration by choosing appropriate steps. The NMF algorithm is described in Algorithm 1.

Algorithm 1. NMF algorithm.

Initialize $W_{ij}\geq 0, H_{ij}\geq 0, \forall i,j$.
for $k=1,2,\ldots$ until convergence **do**
$H_{ij}\leftarrow H_{ij}\frac{(W^TV)_{ij}}{(W^TWH)_{ij}}$,
$W_{ij}\leftarrow W_{ij}\frac{(VH^T)_{ij}}{(WHH^T)_{ij}}$,
$W_{ij}\leftarrow \frac{W_{ij}}{\sum_k W_{kj}}$.
end for

Comparing with other subspace methods with holistic components, such as PCA and LDA, NMF can extract localized components which offer advantages in object recognition, including stability to local deformations, lighting variations, and partial occlusion [3]. In face recognition, NMF has shown to be superior to PCA and LDA [18]. Thus, NMF has been widely investigated recently. In real world, many data exhibit nonlinear structure, however, the linear NMF method cannot learn the nonlinear relations between the data. Therefore, nonlinear NMF should be developed. Two kernel-based nonlinear NMF algorithms have been proposed and are introduced as follows.

2.2. Polynomial NMF (PNMF)

PNMF [12] is the variant of NMF in kernel feature space, aiming at representing the images in a nonlinear way. Each image is firstly embedded in a polynomial feature space P via a polynomial kernel-induced nonlinear mapping

$$\phi: x\in\Gamma\mapsto\phi(x)\in P. \tag{6}$$

The dot product in P can be written by means of polynomial kernel $\langle\phi(x),\phi(y)\rangle=k(x,y)=(x^Ty)^d$, where d is an integer.

The idea of PNMF is to find a set of bases W^ϕ in P to approximate the embedded images, i.e., $\phi(x_i)\approx W^\phi h_i, i=1,2,\ldots,n$. The problem is formulated as minimizing the reconstruction error in P:

$$E_{PNMF}(W^\phi,H)=\|V^\phi-W^\phi H\|_F^2 \quad \text{s.t. } w_i,H\geq 0, i=1,\ldots,r, \tag{7}$$

where $V^\phi=[\phi(x_1),\ldots,\phi(x_n)], W^\phi=[\phi(w_1),\ldots,\phi(w_n)]$. Vectors w_i are called the pre-images of the bases. The detailed algorithm to solve problem (7) is referred to [12].

2.3. Projected gradient kernel NMF (PGKNMF)

PGKNMF [13] is developed to remedy the limitations of PNMF: (1) PNMF cannot guarantee that the limit point is a stationary point, (2) Only polynomial kernel can be used. It solves problem (7) by successively optimizing two subproblems:

$$E_{PNMF}(W^\phi) \quad \text{s.t. } w_i\geq 0, i=1,\ldots,r, \text{ with } H \text{ fixed} \tag{8}$$

and

$$E_{PNMF}(H) \quad \text{s.t. } H\geq 0 \text{ with } W^\phi \text{ fixed}. \tag{9}$$

[2] Semi-NMF decomposes a matrix into a mixed-sign bases and a nonnegative coefficients.