ARTICLE IN PRESS
JID: NEUCOM [m5G;March 5, 2018;21:47]
Neurocomputing 000 (2018) 1–11
Contents lists available at ScienceDirect
Neurocomputing
journal homepage: www.elsevier.com/locate/neucom

Subspace clustering guided convex nonnegative matrix factorization

Guosheng Cui[a,b], Xuelong Li[a], Yongsheng Dong[a,*]

[a] The Center for OPTical IMagery Analysis and Learning (OPTIMAL), Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, Shaanxi 710119, PRChina
[b] University of Chinese Academy of Sciences, 19A Yuquanlu, Beijing, 100049 P R China

ARTICLE INFO

Article history:
Received 13 December 2017
Revised 5 February 2018
Accepted 21 February 2018
Available online xxx

Communicated by Dr. Nianyin Zeng

Keywords:
Convex nonnegative matrix factorization
Subspace clustering
Multiple centroids
Geometry structure
Image clustering

ABSTRACT

As one of the most important information of the data, the geometry structure information is usually modeled by a similarity graph to enforce the effectiveness of nonnegative matrix factorization (NMF). However, pairwise distance based graph is sensitive to noise and can not capture the subspace structure of the data. Reconstruction coefficients based graph can capture the subspace structure of the data, but the procedure of building the representation based graph is usually independent to the framework of NMF. To address this issue, a novel subspace clustering guided convex nonnegative matrix factorization (SC-CNMF) is proposed. In this NMF framework, the nonnegative subspace clustering is incorporated to learning the representation based graph, and meanwhile, a convex nonnegative matrix factorization is also updated simultaneously. To tackle the noise influence of the dataset, only k largest entries of each representation are kept in the subspace clustering. To capture the complicated geometry structure of the data, multiple centroids are also introduced to describe each cluster. Additionally, a row constraint is used to remove the relevance among the rows of the encoding matrix, which can help to improve the clustering performance of the proposed model. For the proposed NMF framework, two different objective functions with different optimizing schemes are designed. Image clustering experiments are conducted to demonstrate the effectiveness of the proposed methods on several datasets and compared with some related works based on NMF together with k-means clustering method and PCA as baseline.

1. Introduction

As one feature extraction technique [1,2], nonnegative matrix factorization (NMF) has became more and more popular in the fields of the computer vision and pattern recognition thanks to the pioneering work of Lee and Seung [3]. In their works, they point out that the nonnegative constraints on the component matrices can automatically lead to the parts-based representation of the data which is closely related to the perception mechanism. Besides this finding, a simple yet effective algorithmic procedure is another contribution of their work. Due to these advantages of NMF, the research around the original NMF and its variants is becoming increasingly flourishing [4–9].

The original NMF method tries to factorize the original nonnegative data matrix into two nonnegative factorial matrices whose product can approximate the original data matrix. To improve the sparseness of NMF, Hoyer [10] has proposed nonnegative sparse coding (NSC) in which a ℓ_1-norm sparse constraint is imposed on the encoding matrix. Localized NMF (LNMF) [11] imposes some local constraints on the factor matrices to help learning a more localized features of the data. Both these two methods can ensure a parts-based representation. For unsupervised variants of NMF, geometry information is another important information that is frequently used to reinforce the effectiveness of NMF methods. Graph regularized NMF (GNMF) [12] first considers to improve the performance of NMF from the geometric perspective. An affinity graph is constructed to encode the local data distributing structure. With this information, the data local geometry structure in the original space can be retained in the learned low dimension space. Neighborbood preserving NMF (NPNMF) [13] tries to encode the local geometry structure with the coefficients of the k nearest neighbors. Although, NPNMF does not use the Heat Kernel to measure the similarity of the nearest neighbors. But the selection of the k nearest neighbors is still based on the Euclidean distance. In order to learn a discriminative and sparse representation, Ren et al. [14] add a rank constraint into the framework of NMF and proposed NMF with regularizations (RNMF). Nie et al. propose a semi-NMF named robust manifold NMF (RMNMF) [15] to improve the robustness of NMF. RMNMF uses $\ell_{2,1}$-norm to measure the residual of the approximation instead F-norm which is sensitive to

* Corresponding author.
E-mail addresses: guosheng.cui.opt@gmail.com (G. Cui), xuelong_li@opt.ac.cn (X. Li), dongyongsheng98@163.com (Y. Dong).

https://doi.org/10.1016/j.neucom.2018.02.067
0925-2312/

Please cite this article as: G. Cui et al., Subspace clustering guided convex nonnegative matrix factorization, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2018.02.067

outliers. Both RNMF and RMNMF have promoted the performance of NMF from different views, but they have one common regularizer in their object function, that is graph Laplacian regularizer. Actually, graph Laplacian regularizer is widely used in various NMF frameworks since its first utilization in GNMF.

For supervised NMF methods, label provided by the data is used to encode the discriminative structure of the data. Discriminative NMF (DNMF) [16] incorporates the Fisher's criterion, which is defined on the encoding matrix, into the framework of NMF directly. An et al. propose a method named manifold-respecting discriminant NMF (NMF-kNN) [17] in which two graph Laplacian regularizers are included. Penalty graph is designed to describe the distinctness among the different classes and intrinsic graph is used to encode the local data distributing structure within the same class. Manifold regularized discriminative NMF (MDNMF) [18] also uses these two graphs to capture the discriminative information of the data. The difference is that, NMF-kNN uses the difference of these two regularizers and MDNMF uses the ratio of these two regularizers. Constrained NMF (CNMF) [19] incorporates the label constraint matrix into the cost function of NMF directly. Graph regularized discriminant NMF (GDNMF) [20] approximates the label indicating matrix with the product of the encoding matrix and a random matrix. In addition, a Laplacian graph is also constructed using label information in this method.

Considering that a Laplacian graph is constructed by a similarity matrix, it can also be called a similarity graph, and used for capturing latent structure information of the data. Generally speaking, there are two ways to build a similarity graph: one way is based on pairwise distance (e.g. Euclidean distance), another way is based on reconstruction coefficients (e.g. sparse representation). Pairwise distance based graph is usually constructed using the Euclidean distance that fails to explore the multi-subspaces structure of the data. However, reconstruction coefficients based graph can be used to capture multi-subspaces structure of the data when building it by using subspace representation coefficients. In this paper, we propose a novel subspace clustering guided convex nonnegative matrix factorization (SC-CNMF). SC-CNMF uses nonnegative subspace clustering to guide convex nonnegative matrix factorization. In this framework, the learning of reconstruction coefficients based graph and the convex nonnegative matrix factorization can be implemented simultaneously. To our knowledge, this kind of unified framework has not been proposed before.

The contributions of the proposed model are listed as follow:

- A unified NMF framework named subspace clustering guided convex nonnegative matrix factorization (SC-CNMF) is proposed. In this framework, nonnegative subspace clustering term, which can capture the multi-subspaces structure of the data, is incorporated to guide the learning of convex NMF. The nonnegative constraint that is imposed on the subspace clustering facilitates the optimizing of the proposed unified framework.
- A local subspace constraint is imposed on nonnegative subspace clustering term to improve the robustness of the proposed model. This constraint is that only s largest values are kept in each column of the learned representation matrix that will be used to construct the similarity matrix in each iteration. The two different ways of using this constraint lead to two different implementations of the proposed model, SC-CNMF$_1$ and SC-CNMF$_2$. And two slightly different optimizing schemes are designed for these two methods.
- Image clustering experiments are conducted on six image datasets in which the proposed methods are compared with several related NMF methods together with k-means clustering method and PCA as baseline. The experimental results reveal the effectiveness of the proposed methods.

The rest of paper is organized as follow: in Section 2, some related NMF methods are briefly reviewed in Section 2.1, then subspace clustering is briefly introduced in the Section 2.2. In Section 3, the details of the proposed model and its two implementations together with their optimizing schemes are described. The experimental results and analysis are in Section 4. The conclusion of this paper is made in Section 5.

2. Reviews of some related works

2.1. NMF, GNMF and convex NMF

Nonnegative matrix factorization (NMF) [3] tries to find two nonnegative factorial matrices, $U \in \mathbb{R}_+^{m\times k}$ and $V \in \mathbb{R}_+^{k\times n}$, whose product can approximate the original data matrix $X = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}_+^{m\times n}$. The objective function of the standard NMF with F-norm is like follow:

$$D_{NMF} = ||X - UV||_F^2. \tag{1}$$

In this equation, we can see that, each original data point can be represented with the linear combination of the column vectors in U weighted by the corresponding column vector in V. So matrix U can be regarded as a basis matrix and matrix V is an encoding matrix. Eq. (1) is non-convex for U and V. But when one of these two matrices is fixed, Eq. (1) is convex for the rest one matrix. Therefore this equation can be minimized with an iterative updating strategy. The nonnegative constraints imposed on U and V cause that the updating rules only allow additive operations. To this end, the following updating rules are obtained:

$$U = U \odot \frac{XV^T}{UVV^T},$$
$$V = V \odot \frac{U^T X}{U^T UV}, \tag{2}$$

where $\odot$ denotes the element-wise multiplication and the division used in above equations are also element-wise.

To make use of the latent embedding manifold structure of the data, Cai et al. [12] incorporate a Laplacian regularizer into the framework of the standard NMF method and propose graph regularized NMF (GNMF). To capture the manifold structure of the data in the original space, a Laplacian graph is constructed which can describe the local geometry structure of the data. The local near neighbor relationship in the original space, i.e. X, can be kept in the learned low-dimension space, i.e. V, by minimizing the graph Laplacian regularizer. With this regularizer, the objective function of GNMF is as follow:

$$D_{GNMF} = ||X - UV||_F^2 + \alpha tr(VLV^T), \tag{3}$$

where α is a parameter to control the influence of the Laplacian regularizer.

Convex NMF [21] constraints each column of the basis matrix to be a convex combination of the data points for the interpretability reason. In Convex NMF, the basis matrix can be denoted as $U = XG$, then its objective function can be written as

$$D_{convexNMF} = ||X - XGV||_F^2. \tag{4}$$

The advantage of this convex constraint is that each column of U can be interpreted as a weighted sum of certain data samples.

2.2. Subspace clustering

Subspace clustering can find out latent subspace structure of the data. Generally, subspace clustering can be divided into three categories: algebraic algorithms, statistical methods and spectral clustering based methods. Among all these three kind of subspace clustering methods, the performance of spectral clustering based