Neurocomputing 173 (2016) 233–244

Contents lists available at ScienceDirect

Neurocomputing

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

Graph regularized and sparse nonnegative matrix factorization with hard constraints for data representation

Fuming Sun *, Meixiang Xu, Xuekao Hu, Xiaojun Jiang

Liaoning University of Technology, Jinzhou 121001, PR China

ARTICLE INFO

Article history:
Received 1 July 2014
Received in revised form
11 October 2014
Accepted 22 January 2015
Available online 5 September 2015

Keywords:
Nonnegative matrix factorization
Graph-based regularizer
Sparseness constraints
Label information

ABSTRACT

Nonnegative Matrix Factorization (NMF) as a popular technique for finding parts-based, linear representations of nonnegative data has been successfully applied in a wide range of applications. This is because it can provide components with physical meaning and interpretations, which is consistent with the psychological intuition of combining parts to form whole. For practical classification tasks, NMF ignores both the local geometry of data and the discriminative information of different classes. In addition, existing research results demonstrate that leveraging sparseness can greatly enhance the ability of the learning parts. Motivated by these advances aforementioned, we propose a novel matrix decomposition algorithm, called Graph regularized and Sparse Non-negative Matrix Factorization with hard Constraints (GSNMFC). It attempts to find a compact representation of the data so that further learning tasks can be facilitated. The proposed GSNMFC jointly incorporates a graph regularizer and hard prior label information as well as sparseness constraint as additional conditions to uncover the intrinsic geometrical and discriminative structures of the data space. The corresponding update solutions and the convergence proofs for the optimization problem are also given in detail. Experimental results demonstrate the effectiveness of our algorithm in comparison to the state-of-the-art approaches through a set of evaluations.

1. Introduction

A fundamental problem in a variety of data analysis tasks is to find an appropriate representation for the given data. The purpose of data representation is to effectively uncover the latent structure of the data so that further learning tasks, such as clustering and classification, can be facilitated. Matrix factorization techniques as fundamental tools for such data representation have been receiving more and more attention. Generally speaking, matrix factorization is non-unique and by far many different methods of doing so have been proposed by incorporating different constraints with different criteria. The canonical techniques include Non-negative Matrix Factorization (NMF) [1,2], the QR Decomposition (QRD), Singular Value Decomposition (SVD), Principal Component Analysis (PCA), Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA), Regularized LDA [3], Deterministic Column-Based Matrix Decomposition [4], etc. Matrix factorization aims to find two or more matrix factors whose product is a good approximation to the original matrix. In practical applications, the dimension of the decomposed matrix factors is often much smaller than that of the original matrix, leading to compact representation of the data points, which is helpful to other learning tasks like clustering and classification. Roughly speaking, matrix decomposition algorithms with strong performance tend to satisfy two basic conditions: (1) it could uncover the intrinsic geometric structures of the data clearly; and (2) it could reduce dimensionality of original data, which in turn can facilitate other learning tasks.

The aforementioned PCA and SVD decompose factorize the matrix as the linear combination of principle components. Unlike these methods, NMF [1,2] specializes in that it enforces the constraint that the elements of the factor matrices must be nonnegative, i.e., all elements must be equal to or greater than zero. And such a nonnegative constraint leads NMF to a parts-based representation of the object in the sense that it only allows additive, not subtractive, combination of the original data. Therefore, it is an ideal dimensionality reduction algorithm for image processing, face recognition, and document clustering, where it is natural to consider the object as a combination of parts to form a whole representation. Since NMF is an unsupervised learning algorithm, it is inapplicable to many real-world problems where limited knowledge from domain experts is available.

After many years of research and development, a plenty of improved methods have been proposed on the basis of NMF. Hoyer

* Corresponding author.
E-mail address: sunwenfriend@hotmail.com (F. Sun).

http://dx.doi.org/10.1016/j.neucom.2015.01.103
0925-2312/

[5] and Dueck [6] et al. computed sparse matrix factorization, in which sparseness constraint is enforced to enhance the ability of learning parts. Li et al. [7] put forward the Local Nonnegative Matrix Factorization (LNMF) method, for learning spatially localized, parts-based subspace representation of visual patterns, giving a set of bases which not only allows a part-based representation of images but also manifests localized features. But it has been pointed out that LNMF could not represent the data very well [8]. Hoyer applied NMF to the sparse code, and proposed the Nonnegative Sparse Coding (NSC) [9]. Furthermore, he brought up a Sparse NMF (SNMF) algorithm which can be controlled explicitly [10]. Cai et al. [11] presented the Graph Regularized Nonnegative Matrix Factorization (GRNMF) approach to encode the geometrical information of the data space by constructing a nearest neighbor graph to model the local manifold structure. When label information is available, it can be naturally incorporated into the graph structure. Concretely, if two data points share the same label, a large weight will be assigned to the edge connecting them. And if two data points have different labels, the corresponding weight is set to be zero. By taking the label information as additional constraints, Liu et al. [12] proposed the Constrained Nonnegative Matrix Factorization (CNMF), which incorporates the label information as additional constraints. Ding et al. [13] proposed a semi-nonnegative matrix factorization algorithm where only one matrix factor is restricted to contain nonnegative entries, while it relax the constraint on the basis vectors. Yuan et al. proposed Binary Sparse Nonnegative Matrix Factorization in [14], making full use of the sparseness property of the basis vector to remove easy-excluded Haar-like box functions.

Due to the effectiveness and importance of data representations, recently variants of NMF have been widely applied in extensive domains, such as document clustering [15,16], audio-visual document structuring [17], speech and image cryptosystems [18], image classification and annotation [19], blind source separation [20], facial expression recognition [21], and image search reranking [22], etc. A comprehensive review about the principles, basic models, properties, and algorithms of NMF is systematically introduced in [23]. What's more, sparseness constraints together with many learning methods have been applied to video search reranking [24] and image categorization [25].

Motivated by recent progress in matrix factorization, we proposed a novel NMF method for data representation by exploiting three constraint conditions, including graph-based regularizer, sparseness requirement and prior label information offered by few labeled data points. The proposed NMF is referred as Graph regularized and Sparse Nonnegative Matrix Factorization with hard Constraints (GSNMFC) to represent the data in a more reasonable way. In this method, we incorporated hard prior label information into the graph to encode the intrinsic geometrical and discriminative structures of the data space, and also take the label information as additional constraints to decompose matrix. Furthermore, we exploited extra sparseness constraints to make the coefficients much sparser. It is under the assumption that if the sparseness levels of the factors are improved, the ability of the learning parts can be enhanced. By combining the three constraints, we expect that further learning performance, such as the recognition rate and clustering results, can be further improved in the new data representation. What's more, we proved the convergence of the raised method. Finally, we carried out the extensive experiments on the common face databases to validate the effectiveness and efficiency of the novel matrix factorization method proposed in this paper.

The main contributions of our work can be summarized as follows:

(1) The proposed method possesses the merit of CNMF, which takes the label information as additional hard constraints and is parameter free. On the other hand, the proposed method has the advantage of GRNMF, which exploits the intrinsic geometric structure of the data distribution and incorporates it as an extra regularization term. In a nutshell, the algorithm presented here incorporates the virtues of the two methods mentioned above. Moreover, the proposed method can have better performance on clustering accuracy and normalized mutual information.
(2) Note that with the sparseness levels of the factors improved the discrimination ability of the learning parts can be enhanced, in which sense sparseness is rather an indispensible constraint to NMF instead of an optional one. Thereby on the basis of GRNMF and CNMF, we enforces additional sparseness constraints, which is beneficial to give rise to a much sparser data that is conductive to other learning tasks like classification and clustering.

The rest of the paper is organized as follows. First, we give brief reviews of related methods including NMF, CNMF and GNMF in Section 2. Next, in Section 3 our proposed algorithm is described in details and theoretical proof of convergence of our optimization scheme is provided as well. Then Section 4 reports extensive experimental results and corresponding analyses on two popular datasets followed by some concluding remarks made in Section 5.

2. Brief reviews

2.1. NMF

Given a data matrix $X = [x_1, x_2, \dots, x_n] \in R^{m\times n}$, where $x_i \in R^m$ is a sample vector, and the elements of each sample vector is nonnegative. The goal of NMF is to seek two nonnegative matrices $U \in R^{m\times k}$ and $V \in R^{n\times k}(k \le mn/(m+n))$. Especially, the elements of the two matrices are all negative, in order to guarantee that the similarity between X and UV^T is the highest, that is to say, it reduces to minimize the following objective function:

$$O_F = \|X - UV\|^2 \quad \text{s. t.} \quad U > 0, V > 0 \tag{1}$$

where O_F is referred as the Frobenius norm.

Eq. (1) is a description form of the objective functions from the Euclidean space perspective. And there is another description form from the point view of divergence as below:

$$O_{KL} = D\left(X \| UV^T\right) = \sum_{i,j}\left(x_{ij}\log\frac{x_{ij}}{(uv^T)_{ij}} - x_{ij} + \left(uv^T\right)_{ij}\right) \tag{2}$$

It can be proven that (1) and (2) are both convergent. According to the aforementioned equations, the multiplicative iterative updating formula [1,2] can be achieved as below.

$$u_{ik} \leftarrow u_{ik}\frac{(XV)_{ik}}{(UV^TV)_{ik}} \tag{3}$$

$$v_{jk} \leftarrow v_{jk}\frac{\left(X^TV\right)_{jk}}{\left(V^TVU^T\right)_{jk}} \tag{4}$$

where $U = [w_{ik}]$, $V = [v_{jk}]$. At the very beginning of the iterative update process, the two nonnegative matrices U_o and V_o are initialized at random. The iterative update procedure is executed repeatedly according to (3) and (4) until the given terminal condition is met. Ultimately, the final U and V can be obtained.