Neurocomputing 173 (2016) 224–232

Contents lists available at ScienceDirect

Neurocomputing

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

General subspace constrained non-negative matrix factorization for data representation

Yong Liu [a,b,*], Yiyi Liao [b], Liang Tang [c], Feng Tang [a], Weicong Liu [b]

[a] State Key Lab of Industrial Technology Control Technology, Zhejiang University, Hangzhou 310027, China
[b] Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou 310027, China
[c] China Ship Development and Design Center, Wuhan 430064, China

ARTICLE INFO

Article history:
Received 14 June 2014
Received in revised form
10 November 2014
Accepted 16 November 2014
Available online 3 September 2015

Keywords:
Non-negative matrix factorization
Subspace
Generalize framework

ABSTRACT

Nonnegative matrix factorization (NMF) has been proved to be a powerful data representation method, and has shown success in applications such as data representation and document clustering. However, the non-negative constraint alone is not able to capture the underlying properties of the data. In this paper, we present a framework to enforce general subspace constraints into NMF by augmenting the original objective function with two additional terms. One on constraints of the basis, the other on preserving the structural properties of the original data. This framework is general as it can be used to regularize NMF with a wide variety of subspace constraints that can be formulated into a certain form such as PCA, Fisher LDA and LPP. In addition, we present an iterative optimization algorithm to solve the general subspace constrained non-negative matrix factorization (GSC NMF). We show that the resulting subspace has enriched representation power as shown in our experiments.

1. Introduction

Finding a suitable representation is a fundamental problem in many machine learning tasks, such as pattern recognition and object detection [1–7]. A good representation can capture the underlying structure of the data and can reduce the dimensionality so as to make the higher level inference easier. Subspace representations construct a subspace from the original high dimensional space and represent them as the projection on the subspace. It has been shown that subspace methods not only reduces the computational cost due to the lower dimensionality but also makes the higher level inference easier.

Subspace methods such as principal component analysis (PCA) [8,9,1], linear discriminative analysis (LDA) [10,11] and locality preserving projection (LPP) [12] can be understood as matrix factorization subject to different constraints. These constraints are usually designed to find basis functions satisfying certain properties. Principal components analysis enforces an orthogonality constraint of the basis vectors, resulting in an orthogonal subspace to capture the major variance of the data. As a well-known dimension reduction method, PCA is extended in different ways, such as incremental learning and tensor analysis [13–15]. Extension approaches of LDA and LPP are also proposed for performance improvement [16–19]. However, the resulting basis and coefficient vectors can be negative, which does not have intuitive psychological interpretation. Non-negative Matrix Factorization (NMF) is a subspace method with nonnegative constraints on both the basis and coefficients. The non-negative constraints lead to a parts-based representation because they allow only additive, not subtractive combinations. Such a representation encodes the data using few active components, which makes the basis easy to interpret. The previous research works have shown the superior performance of NMF on document clustering [20], text mining [21,22], pattern recognition [23,24] and audio analysis [25,26].

However, the non-negative constraints alone may not be enough to capture the underlying structure of the data as other subspace methods for example PCA do. In this paper, we present a framework to enforce general subspace constraints into NMF. This framework is general as it can be used to regularize NMF with a wide variety of subspace constraints that can be formulated into a certain form such as PCA, LDA and LPP. In addition, we present an iterative optimization algorithm to solve the general subspace constrained NMF. We show that the resulting subspace has enriched representation power as shown in our experiments.

There are also some other work that tries to incorporate constraints into the NMF. Local non-negative matrix factorization (LNMF) [27] has been proposed to achieve a more localized NMF algorithm with the aim of computing spatially localized basis adding orthogonality constraints that modify the objective function.

* Corresponding author at: State Key Lab of Industrial Technology Control Technology, Zhejiang University, Hangzhou 310027, China
E-mail address: yongliu@iipc.zju.edu.cn (Y. Liu).

http://dx.doi.org/10.1016/j.neucom.2014.11.099
0925-2312/

Some similar works focus on constraining the orthogonality such as [28] and [29]. The former solves the optimization problem with the orthogonality constraints, while the later embeds the constraints as part of the cost function. In Sparse NMF [30], the author enforces the sparseness constraints explicitly in the objective function. However, these algorithms and their solutions are specifically designed for a particular constraint, which are in contrast with our approach since we aim at providing a general theoretical framework and solution.

The rest of the paper is organized as follows: Section 2 gives a brief review of the NMF. The general theoretical framework for NMF with subspace constraints and their three examples PCA NMF, Fisher NMF and LPP NMF are presented in Section 3. The optimization algorithm is discussed in Section 4. The experimental results will be shown in Section 5 and we conclude the paper in Section 6.

2. A brief review of NMF

Generally, NMF [31] can be presented as the following optimization problem:

$$\min_{\mathbf{B},\mathbf{H}} C(\mathbf{X} \approx \mathbf{B}\mathbf{H}), \quad \text{s. t.} \quad \mathbf{B}, \mathbf{H} \ge 0$$

Here, $\mathbf{X} = [x_{ij}] \in \mathcal{R}^{d\times n}$, each column of $\mathbf{X}$ is a sample vector. NMF aims to find two non-negative matrices $\mathbf{B} = [b_{ij}] \in \mathcal{R}^{d\times r}$ and $\mathbf{H} = [h_{ij}] \in \mathcal{R}^{r\times n}$ whose product can well approximate the original matrix $\mathbf{X}$. $C(\cdot)$ denotes the cost function. There are normally two kinds of cost functions to represent the approximation in the NMF optimization. Let $\mathbf{Y}=\mathbf{B}\mathbf{H}$, the first cost function is the Euclidean distance between two matrices:

$$C_1 = \|\mathbf{X} - \mathbf{Y}\|^2 = \sum_{ij} \left(x_{ij} - y_{ij}\right)^2$$

The second cost function is the K–L divergence between two matrices:

$$C_2 = D(\mathbf{X} \,\|\, \mathbf{Y}) = \sum_{ij} \left(x_{ij} \log \frac{x_{ij}}{y_{ij}} - x_{ij} + y_{ij} \right)$$

Although the C_1 and C_2 are convex in $\mathbf{B}$ only or $\mathbf{H}$ only, they are not convex in both variables together. Thus Lee and Seung [32] presented iterative update algorithms to find the local minima of the objective function C_1 and C_2 [32].

3. General subspace constrained NMF

In this section, we present a general subspace constrained non-negative matrix factorization (GSC NMF) framework, which can enforce various subspace constraints into NMF. Let $\mathbf{U} = [u_{ij}] = \mathbf{B}^{\mathbf{T}}\mathbf{B}$, the problem is formulated as follows:

$$O_U = C(\mathbf{X} \approx \mathbf{B}\mathbf{H}) + \alpha \sum_{i,j} u_{ij} - \beta\, \mathrm{Tr}(\mathbf{H}\mathbf{L}\mathbf{H}^{\mathbf{T}}) \tag{1}$$

α, β are const real number, and $\mathbf{L} \in \mathcal{R}^{n\times n}$ is the parameter matrix. The cost function C is either C_1 or C_2. When $\alpha > 0$, minimizing $\sum_{ij} u_{ij}$ leads to the basis (b_i), which are orthogonal [33,27]. Table 1 shows different parameter settings for the GSC NMF and their corresponding subspace constrains. With page limits, we will only introduce the PCA NMF, Fisher NMF and LPP NMF in detail in the following sections.

Table 1
Various subspace constrained NMF via different parameter settings.

	C	α	β	$\mathbf{L}$
LNMF [27,33]	C_1 or C_2	$\alpha > 0$	$\beta > 0$	$\mathbf{I}$
PCA NMF	C_1 or C_2	$\alpha > 0$	$\beta > 0$	$\mathbf{L} = \frac{1}{n}\mathbf{I} - \frac{1}{n^2}\mathbf{e}\mathbf{e}^{\mathrm{T}}$
Fisher NMF	C_1 or C_2	$\alpha = 0$	$\beta < 0$	$\mathbf{L} = \mathbf{I} - 2\mathbf{W} + \frac{1}{n}\mathbf{e}\mathbf{e}^{T}$
LPP NMF	C_1 or C_2	$\alpha = 0$	$\beta < 0$	$\mathbf{L} = \mathbf{D} - \mathbf{S}$

3.1. PCA NMF

The main idea of classical PCA is trying to maximize the representation vectors' variance while keeping the orthogonality of the basis. Assuming that the sample data set is $\{x_1, x_2, \ldots, x_n\}$, and the linear transform for PCA can be denoted as $b^T x_i = y_i$, here b is the basis vector of $\mathbf{B}$ and y_i are the representation vector. Then the optimization function of PCA can be denoted as

$$\max_{\mathbf{b}} \sum_{i=1}^{n} (y_i - \bar{y})^2, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

The vectors of $b_1, b_2, \ldots, b_r$ are orthogonal.

When concerning the NMF form of $\mathbf{X} \approx \mathbf{B}\mathbf{H}$, we have $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$, $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_r]$ and $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n]$. Then the column vector of $\mathbf{H}$ can be viewed as the projection of original data set $\mathbf{X}$ in the subspace constructing with the column vectors of $\mathbf{B}$, thus $\mathbf{x}_i \approx \mathbf{B}\mathbf{h}_i$.

We then let $\mathbf{L} = (1/n)\mathbf{I} - (1/n^2)\mathbf{e}\mathbf{e}^{\mathrm{T}}$. $\mathbf{I}$ is the identity matrix with order of n and $\mathbf{e}$ is the n dimensional vector with all the elements equaling to 1. We use $\mathbf{m}$ to denote the mean of the project vectors, that is $\mathbf{m} = (1/n)\sum_{i=1}^{n} \mathbf{h}_i$, and then

$$\begin{aligned}
\mathbf{H}\mathbf{L}\mathbf{H}^{\mathrm{T}} &= \frac{1}{n}\mathbf{H}(\mathbf{I} - \frac{1}{n}\mathbf{e}\mathbf{e}^{\mathrm{T}})\mathbf{H}^{\mathrm{T}} \\
&= \frac{1}{n}\mathbf{H}\mathbf{H}^{\mathrm{T}} - \frac{1}{n^2}(\mathbf{H}\mathbf{e})(\mathbf{H}\mathbf{e})^{\mathrm{T}} \\
&= \frac{1}{n}\sum_i \mathbf{h}_i\mathbf{h}_i^{\mathrm{T}} - \frac{1}{n^2}(n\mathbf{m})(n\mathbf{m})^{\mathrm{T}} \\
&= \frac{1}{n}\sum_i (\mathbf{h}_i - \mathbf{m})(\mathbf{h}_i - \mathbf{m})^{\mathrm{T}} \\
&\quad + \frac{1}{n}\sum_i \mathbf{h}_i\mathbf{m}^{\mathrm{T}} + \frac{1}{n}\sum_i \mathbf{m}\mathbf{h}_i^{\mathrm{T}} - \frac{1}{n}\sum_i \mathbf{m}\mathbf{m}^{\mathrm{T}} - \mathbf{m}\mathbf{m}^{\mathrm{T}} \\
&= E[(\mathbf{h} - \mathbf{m})(\mathbf{h} - \mathbf{m})^{\mathrm{T}}] + 2\mathbf{m}\mathbf{m}^{\mathrm{T}} - 2\mathbf{m}\mathbf{m}^{\mathrm{T}} \\
&= E[(\mathbf{h} - \mathbf{m})(\mathbf{h} - \mathbf{m})^{\mathrm{T}}].
\end{aligned}$$

Here, $E[(\mathbf{h} - \mathbf{m})(\mathbf{h} - \mathbf{m})^{\mathrm{T}}]$ is the covariance matrix of the projections and thus maximizing $\mathbf{H}\mathbf{L}\mathbf{H}^{\mathrm{T}}$ is equivalent to maximizing $\sum_{i=1}^{n} \|\mathbf{h}_i - \mathbf{m}\|$, which is the core optimization function of PCA. At the same time, minimizing $\sum_{i\neq j} u_{ij}$ will guarantee that all the basis vectors are orthogonal.

Then let $\alpha, \beta > 0$ and $\mathbf{L} = (1/n)\mathbf{I} - (1/n^2)\mathbf{e}\mathbf{e}^{\mathrm{T}}$, the optimization function (1) is a PCA constrained NMF.