Journal of Visual Communication and Image Representation 55 (2018) 477–488

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

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

Non-negative matrix factorization via discriminative label embedding for pattern classification☆

Wenjie Zhu*, Yunhui Yan

School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China

ARTICLE INFO

Keywords:
Non-negative matrix factorization (NMF)
Discriminative label embedding
Orthogonality constraint
Pattern classification

ABSTRACT

As one of the most commonly used dimension reduction approaches, discriminant non-negative matrix factorization (NMF) has been widely used for data representation in the pattern classification task. However, the previous discriminant NMFs emphasize the Fisher criterion or maximum margin criterion which has high requirement to the distribution of data. Therefore, this work proposes a discriminative label embedded NMF (LENMF) algorithm. LENMF takes into account the discriminative label embedding to obtain the low-dimensional projected data and orthogonal property of the non-negative basis to strength the ability of parts-based representation. Besides, LENMF is extended in the kernel space to explore the nonlinear relations of data. By integrating the non-negative constraint, discriminative label embedding, and the orthogonal property into the proposed objective, the multiplicative updating rules have been given in this work. Experiment results on the challenging face, object, document, and digit databases illustrate the performance of the proposed algorithm.

1. Introduction

There have been numerous applications of data processing introducing sorts of the prior knowledge into the representation learning procedure. Generally, the suitable representation assists in capturing the intrinsic structure of data. For the pattern classification task, it is inclined to extract the distinguishable feature to represent the original data. In the past decades, there have been sorts of methods to extract the feature of data, such as sparse coding [1–3], linear discriminant analysis [4–6], feature selection [7–9], pixel-based feature description [10–13], and deep learning technique [14,15]. These methods devote to learning the coefficient matrix or the projection matrix for minimizing the reconstruction error while boosting the discriminative ability. However, the methods ignore the non-negative property of the related variables. Both the coefficient and basis matrices derived from the mentioned methods have negative entries. Intuitively, the handled digital data, such as image, video, and document, is non-negative and should be represented by the combination of the non-negative sub portions which are expected to indicate the visual or intuitively meaningful parts of the original data. For instance, the face image can be recognized by observing the discriminant parts, such as nose, eyes, and mouth. From this perspective, non-negative property is more consistent with the psychological intuition.

Linear discriminant analysis (LDA) and principal component analysis (PCA) are the famous approaches to dimension reduction [16]. Both LDA and PCA aim at learning a projection, however, they adopt different learning schemes. LDA incorporates the label information into the learning scheme and attempts to learn a projection which can map the data points into a low-dimensional space in which the centers of the different classes have the maximum distances. PCA learns the projection in unsupervised fashion and leads to a low-dimensional space in which data points have the maximum variance. Subsequently, matrix factorization has become popular for data representation. In the real applications of pattern recognition and information retrieval, the original input data is of high dimension which increases the pressure of data processing. Accordingly, the matrix factorization technique aims to decompose the high-dimensional input data matrix into some low-dimensional matrices. Singular value decomposition (SVD) [17], vector quantization (VQ) [18], and non-negative matrix factorization (NMF) [19] are some of the representative matrix factorization techniques. SVD attempts to represent the input matrix in a low-rank approximation, and has been applied to face recognition successfully. VQ maps the input data matrix into binary vectors and is commonly used for information retrieval task, while NMF aims to search for two matrices whose entries are non-negative and product is approximate to constructing the input data matrix. There have been various works

☆ This paper has been recommended for acceptance by Xuelong Li.

* Corresponding author.
E-mail addresses: wenjie_zh@126.com (W. Zhu), yanyh@mail.neu.edu.cn (Y. Yan).

https://doi.org/10.1016/j.jvcir.2018.06.030
Received 29 August 2017; Received in revised form 29 May 2018; Accepted 30 June 2018
Available online 07 July 2018
1047-3203/

indicating that the non-negativity constraint leads to the parts-based representation of the data. In recent years, non-negative property has been used in feature extraction [20–25]. Diverse from the PCA, LDA, VQ, and SVD, NMF only allows additive, not subtractive combination of the input data, and thus it is naturally favor to sparse, parts-based representation which is more robust than non-sparse, global representations.

Generally, the NMF works can be categorized into four classes [26], including sparse NMF [27,28], orthogonal NMF [29–31], manifold NMF [32] and discriminant NMF [33–35]. Sparse NMF focuses the sparseness property of the matrices. The original NMF enforces the matrices to be non-negative. Meanwhile, NMF may extract the sparse information. However, NMF is different from sparse coding. Sparse coding learns a full rank representation basically, whilst NMF pursues for a low-rank representation. Furthermore, NMF is not dictionary learning. Dictionary learning aims to optimize an over-complete basis matrix, however, basis matrix generated by NMF is under-complete. Orthogonal NMF can obtain the parts-based representation to boost the psychological and physiological representation ability. Manifold NMF expects to extend the NMF on the manifold structure data. This work focuses on the discriminant NMF to enhance the discriminant ability of NMF for image classification tasks.

As one of the most famous matrix factorization methods, NMF has been attracting more and more attentions in recent years. Li et al. imposed extra constraints to solve the localized and parts-based decomposition by extending the original NMF and this work is called as LNMF [36]. To obtain sparse encoding vectors, Hoyer incorporated the sparseness constraint with the original NMF and proposed the non-negative sparse coding (NSC) [37]. Employing the data geometric structure, Cai et al. proposed a graph regularized NMF (GNMF) [20]. The geometric structure encoded by a k-nearest-neighbors measurement is usually used in dimension reduction methods. Compared with these unsupervised NMF variants, the label information involved discriminant NMF variants are more feasible for the pattern classification task. As the previous discriminant NMF variant, the Fisher-NMF (FNMF) [38] was proposed to encode discriminant information into NMF. Then, Zafeiriou et al. extended FNMF by adding an extra term of scatter difference to the objective function of NMF to obtain the discriminant subspace. In [33], Zafeiriou et al. extended the original NMF by enforcing not only the spatial locality, but also the separability between classes in a discriminant manner and proposed a discriminant NMF (DNMF) with application to frontal face verification. Based on the DNMF work, Kotsia et al. proposed a projected gradient method for DNMF (PG-DNMF) [34] which is suitable for classification tasks. The maximum margin criterion (MMC) has been successfully used in LDA for avoiding the small sample size problem. Due to the performance of MMC for exploring the discriminant information, there have been MMC-based NMF works [35,39,40]. In [39], the MMC acting on the coefficient matrix is involved into the NMF model for pattern classification tasks. Then, Liu et al. proposed a spatial NMF algorithm based on the max-margin coding in [40]. By employing the projective NMF model, Guan et al. take into account the MMC for discriminant learning and achieve impressive classification results [35].

The latent discriminant information plays a vital role for the image classification task. However, the unsupervised NMF methods lack of extracting enough discriminant information. As for the pattern classification tasks, discriminant NMF variants which utilize the label information are the better choices. However, there are still some issues for the mentioned discriminant NMF works:

1. Both Fisher criterion and MMC are used acting on the coefficient matrix to enforce the discriminant learning for supervised NMF. However, the objective is keeping the balance between the construction and the discriminant terms. It cannot obtain an discriminant representation of the original data by using the basis matrix exactly. Therefore, the constraints which are expected to explore the discriminant information have to cater to the representation of the unseen data[1] for improving the classification performance.
2. The label information measured in the discriminant NMF works is transformed using the distances between the class and another one or among the data points in the same class. This measurement is too rigorous to keep the independence of the data points and relies on the distribution of data largely. In other words, this measurement which employs the mean data points works on the assumption that the global distribution of data points is consistent with the local one. However, this is rigorous for the real data.

To this end, this paper proposes a discriminative label embedded NMF (LENMF) based on the squared Euclidean distance. LENMF integrates the orthogonality constraint and supervised label information for the basis matrix into the objective function. By doing so, the parts-based representation can extract the discriminant information which is consistent with the label. Besides, the kernel function is employed to assist in exploring the nonlinear relations of the data. According to the designed classifier, the proposed LENMF fulfills a novel discriminant NMF method in the pattern classification task.

Most of the contributions of this paper are summarized as follows:

1. The proposed algorithm devotes to searching for the orthogonal basis matrix in the kernel space for the parts-based representation and indicates the discriminant localization.
2. The proposed algorithm enforces the label which indicates the discriminant information to be embedded into learning the projected representation which is robust to the discriminant feature extraction and guides the latent discriminant information of data to the right label.
3. The designed linear classifier increases the effectiveness of the pattern classification task. Comparison experiments on the challenging databases well validate the performance of the proposed algorithm.

The remainder of this paper is organized as follows: Section 2 briefly introduces the NMF variants; Section 3 describes the proposed algorithm in detail; Section 4 presents the experimental results and Section 5 concludes this paper.

2. Brief review of NMF variants

Before specifying the related work, we first introduce the notations used in the whole paper. Given a real $m \times n$ matrix $A \in R^{m\times n}$, A_{ij} denotes the element in the i-th row and j-th column. Then, the i-th row vector and j-th column vector are represented using $A_{i,:} \in R^n$ and $A_{:,j} \in R^m$, respectively. The Frobenius norm of the matrix is denoted as $\|\cdot\|_F$.

2.1. Original NMF

Given an input data matrix $X = \{X_{:,j}\} \in R^{m\times n}$, NMF aims to search for two non-negative matrices $B \in R^{m\times r}$ and $C \in R^{r\times n}$ whose product can approach to X:

$$X \approx BC. \tag{1}$$

In the related work, there are two commonly-used criterions measuring the cost function. The first one is the square of Euclidean distance,

[1] The unseen data means the data does not participate into the training processing. The traditional NMF is a transductive learning algorithm and relatively weak in capturing the non-negative representation of the unseen data.