Journal of Computers Vol. 32 No. 6, 2021, pp. 107-121
doi:10.53106/199115992021123206009

A Survey: Object Feature Analysis Based on Non-negative Matrix Factorization

Shuang Ma[1*], Jinhe Liu[2], Liang Gao[3]

[1] School of Automation, Beijing Information Science and Technology University, Beijing, China
mashuang@bistu.edu.cn

[2] Beijing Aerospace Xinli Technology Co., Ltd, Beijing, China
duanmuyiyi@sina.com

[3] Shenzhen Power Supply Bureau Co., Ltd, Shenzhen, China
349805992@qq.com

Received 30 March 2021; Revised 4 July 2021; Accepted 1 August 2021

Abstract. Non-negative matrix decomposition (NMF) algorithm is the decomposition of all the elements in the matrix under the condition that each element should be non-negative. As a relatively effective technique for dimensionality reduction, NMF has been widely applied in the area of mathematics-physics, engineering and image feature analysis. However, there are few systematic reviews on NMF, especially the application of NMF in image feature extraction. In this paper, NMF algorithms are classified into standard NMF algorithms and improved algorithms according to the theory and application characteristics of different approaches. The basic principles, advantages and shortcomings of these NMF algorithms are systematically analyzed and compared. Firstly, the basic idea of non-negative matrix factorization is introduced, and its application in image feature extraction is illustrated by taking face image as an example. Then, the basic methods and improved algorithms of NMF are emphatically discussed in detail. The examples of local features extracted by different NMF methods are demonstrated on the basis of object feature analysis methods. Finally, the problems to be solved in the practical application of NMF are put forward for improving.

Keywords: non-negative matrix factorization, local feature, sparse coding, two dimensional NMF

1 Introduction

With the development of science and technology, tens of thousands of images with rich context information have been generated every day. These pictures are closely related to our lives, but how to quickly and effectively retrieve, classify and extract valuable features from images has become a hot topic in the research of image processing or other related areas. Image analysis refers to the understanding of properties of each object and the relationship between them in an image. The representation of visual features is the key part of image understanding. However, the "gap" between the low-level visual features and the high-level semantic features has hindered the development of image processing technology. How to find a reasonable way to display data and extract hidden features in data has become an urgent problem.

Features like color, texture, shape (contour, edge) and spatial relationship (position) are always used for object description in traditional image analysis methods. Feature extraction is to decompose different kinds of feature information from the image, mainly including visual features and statistical features. Visual features refer to the shape and color of images with intuitive significance. Statistical features are the statistical calculation of image pixels, textures and other features. The typical methods for statistical

* Corresponding Author

feature extraction are singular value decomposition (SVD), principal component analysis (PCA), vector quantization (VQ), and independent component analysis (ICA). The images obtained by these methods are a global description, and the value of pixels can be positive or negative. There might also be a subtraction relationship between original images. In many cases, the negative value of data has no physical meaning, for example, the sound signal, signal energy, image pixel brightness and so on. Furthermore, the subtraction of data might cause the loss of connection between theoretical models and practical problems [1].

To solve these problems, Lee *et al*. proposed the method of nonnegative matrix decomposition (NMF) in 1999[2]. By decomposing the original data into two low dimensional sub-matrixes, a base matrix and a coefficient matrix are obtained. Each column of the base matrix represents a local feature, and each column of the coefficient matrix shows the representation of a sample in a low dimensional space. Since all data elements are non-negative, the original data can be regarded as a pure additive combination of all base features. This additive operation makes each base feature “sparse” and “localized”. Compared with other methods, NMF can be regarded as a process of forming a whole from local combination, which is more in accord with the perception process of human vision. The sparsity of NMF results enables it to suppress the adverse effects of external changes on feature extraction, such as partial occlusion, illumination changes and object rotation. In addition, it is possible to calculate the low rank approximation of the original matrix by NMF. Thus, NMF also provides an effective way for mining the structural features hidden in the data.

NMF can be solved by simple iterative methods and it is easy to converge. For now, it has been widely applied in the area of face recognition, image clustering, image retrieval, blind source separation, image retrieval and many other areas. It has been several decades since NMF was proposed. The related algorithms have been greatly developed and successfully applied to solve various practical problems. It is necessary to systematically review the existing algorithms, however, the number of such articles is quite few at present. In this paper, the basic concept and improved methods of NMF are first introduced, and then the application of NMF in object feature analysis are analyzed by taking face images for example. We finally summarized the improved algorithm of NMF and the problems to be further studied in the future.

The main contribution of this paper is to classify the existing NMF algorithms into standard NMF algorithms and improved NMF algorithms according to whether non-negative is the only restriction on the decomposition result of NMF model. The standard NMF algorithms mainly include the approach which takes KL divergence as the objective function and the approach which takes Euclidean distance as the objective function. The improved NMF algorithms mainly include sparsity enhanced NMF, discriminative NMF, manifold based NMF and two-dimensional NMF. The remainder of the article is organized as follows: Section II provides the model of standard NMF algorithms. The improved NMF algorithms are discussed in Section III. Finally, conclusions are drawn in Section IV.

2 Standard NMF Algorithms

The traditional image features, such as color [3-4], shape [5-8] and texture [9-11] are still widely used in modern image analysis and understanding. In these models, the description of images is generally presented in the form of statistical data. In fact, there is an inevitable gap between these statistics and human understanding of the image content. Human has the ability to learn from tens of thousands of images in daily life, the understanding of the image content is based on the previously accumulated human experience and knowledge which cannot be expressed by low-level features. How to describe the image content and make it consistent with human understanding is the main emphasis of many researchers. Undoubtedly, the introduction of NMF is an important step in this direction.

The research on non-negative matrices started as early as the 1970s, but it was not noticed at that time. By early 1990s, Paatero, Tapper and other scientists applied non-negative matrix to environmental treatment and satellite remote control [12-13]. In 1997, Lee and Seung proposed an unsupervised learning algorithm with flexible constraints based on cone coding and convex coding [14]. Since then NMF has been widely applied in different areas. The basic idea of NMF can be described as follows: for a given non-negative matrix $V_{n\times m}$, attempt to find two non-negative matrices, $W_{n\times r}$ and $H_{r\times m}$, which satisfies $V \approx WH$ or $V_{ij} \approx (WH)_{ij} = \sum_{k=1}^{r} W_{ik} H_{kj}$. Thus, a non-negative matrix can be decomposed into the product of two non-negative matrices. Matrix W is the base matrix while matrix H is the coefficient matrix. Generally, the value of r should be strictly selected under the constraint $(n+m)r << nm$. Since the matrix before and after decomposition contains only non-negative elements, the column vector of the original matrix V can be interpreted as the weighted sum of all column vectors in the base matrix W. The weight coefficient is the element in the corresponding column vector in H. This representation based on the combination of base vectors has a very intuitive reflection of the concept of “parts constitute the whole” in human mind.

Taking face images as an example, Fig. 1 shows face expressions based on NMF, VQ and PCA, respectively. Obviously, most of the values in the base vectors (column vectors of W) and weighted vectors (row vectors of H) in NMF are zero. It means that the base image matrix W and coded image matrix H are sparse. The base image consists of the mouth, nose, and other facial elements in different positions and forms, and the whole face is created by combining these different parts together. This is significantly different from the holistic face representation in VQ and PCA algorithms.

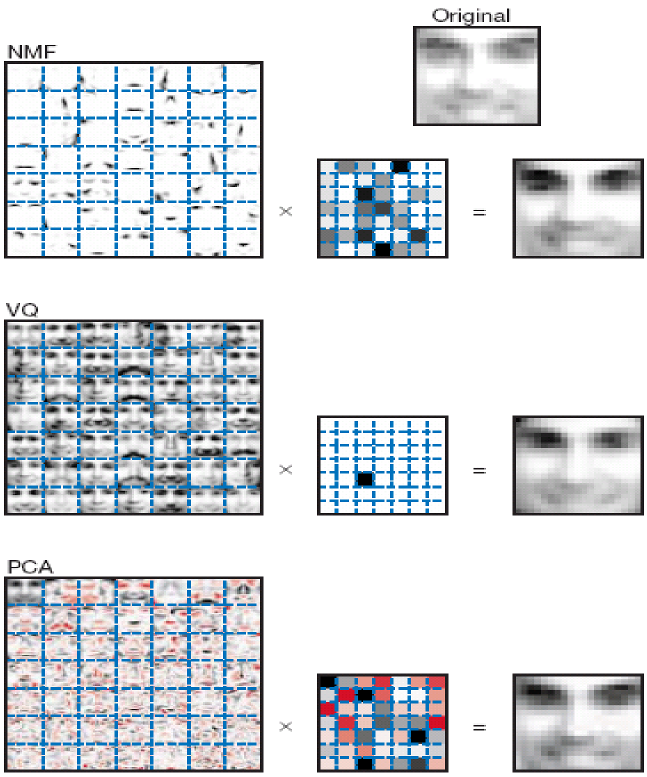

Fig. 1. The comparison of NMF, VQ and PCA in face expression

It is pointed out that NMF is an optimization problem that can be solved by corresponding optimization rules. Donoho *et al*. theoretically analyzed the condition for the existence of unique solution for NMF [15]. This condition tells us that a reasonably constructed objective function is the key to solve NMF as an optimization problem. It is feasible to obtain a local optimal solution of NMF by alternately optimizing matrix W and matrix H.

2.1 Standard NMF Based on the *KL* Divergence

How to define an appropriate objective function to measure the error between matrix V and the product of matrix W and H is critical for NMF. Lee *et al*. [16] proposed an objective function based on the *KL* divergence from information theory. The objective function can be described as:

$$F = \sum_{i,j} (V_{ij} \log \frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij}) \tag{1}$$

The solving methods of NMF mainly includes multiplicative update approach, gradient descent algorithm and alternate least square algorithm. We will describe these three methods in detail.

2.1.1 Multiplicative Update Approach

The basic multiplicative update approach was proposed by Lee and Seung in [16]. The iteration is carried out according to following rules.

$$H_{au} \leftarrow H_{au} \frac{\sum_i W_{ia} V_{iu} /(WH)_{iu}}{\sum_k W_{ka}} \tag{2}$$

$$W_{ia} \leftarrow W_{ia} \frac{\sum_u H_{au} V_{iu} /(WH)_{iu}}{\sum_v H_{av}} \tag{3}$$

The method discussed above was proven to converge to a local minimum point through gradients and continuous non-increment. However, Gonzalez and Zhang [17] pointed out that this approach is technically incorrect because it only presented evidence of continuously declining convergence without considering convergence in saddle point scenarios. Compared with other methods, the standard NMF algorithm proposed in [16] requires more iterative steps and the computation required for each iteration is O(*mnr*). Since NMF is not a global optimization method, sometimes the global optimal solution will not be obtained. To improve the performance of standard NMF, many researchers have proposed different modifications, for example, Gonzalez and Zhang [17] presented a variation of the Lee-Seung algorithm [16] and obtained improved performance. Lin [18] introduced two projected gradient methods for NMF and demonstrated that one of them converges faster than the popular multiplicative update approach.

2.1.2 Gradient Descent Method

In order to improve the convergence speed of the standard NMF algorithm, an optimization strategy similar to the EM algorithm was applied to alternately optimize the *KL* divergence objective function. Thus, the gradient descent method with monotone descent convergence is obtained. The iteration rules of this algorithm are as follows.

$$W^{k+1} \leftarrow W^k - \varepsilon W \nabla_W f(W^k, H^k) \tag{4}$$

$$H^{k+1} \leftarrow H^k - \varepsilon H \nabla_H f(W^k, H^k) \tag{5}$$

εW and εH are step size in formula (4) and (5). $\nabla_W f(W^k, H^k)$ and $\nabla_H f(W^k, H^k)$ are the partial derivatives of W and H in the objective function. This method requires that the direction of each iteration must be non-negative and must also be the direction of the fastest descent. Gradient descent algorithm is an iterative step-size self-learning method, which well mediates the contradiction between algorithm efficiency and implementation simplicity.

The selection of εW and εH is closely related to the convergence of the gradient descent method. In common cases, only simple geometric rules are adopted for the selection of step size, for example, multiply or divide the step size by a factor after each iteration. This operation is sensitive to the initial values and it always leads to unsatisfactory results. Strictly speaking, Lee and Seung's method is also a gradient descent method with a better step selection method but low convergence speed. In [19], the step size is initially set to be one and then halved every iteration. Although this method is simple, the non-negativity of data cannot be guaranteed. The most commonly used method for choosing the appropriate step size in gradient descent method is a simple projection way. The negative value in the matrix is set to be 0 after each iteration to ensure the non-negativity of the data.

2.1.3 Alternating Least Square Algorithm

The objective function of *KL* divergence is convex to matrix *W* or *H*, but not convex to matrix *W* and *H* simultaneously [20]. Therefore, the least square of another matrix can be calculated when one matrix is fixed. The basic alternating least square algorithm can be described as follows.

$$W^{k+1} \leftarrow \arg\min_{W\geq 0} f(W, H^k) \tag{6}$$

$$H^{k+1} \leftarrow \arg\min_{H\geq 0} f(W^{k+1}, H) \tag{7}$$

To ensure non negativity, negative value is still forced to be zero in alternating least square algorithm. In multiplicative update approach, once zero appears in the data set, then it must be zero in the subsequent iteration. This operation significantly increases the sparsity of data. Compared with other methods, the alternating least square algorithm needs less computation.

NMF has been widely used in many areas, but its slow convergence is always a bottleneck. Therefore, many researchers have tried to improve the standard NMF algorithms to get faster convergence. Gonzalez and Zhang [17] proposed an internal point gradient method to accelerate the standard NMF. Lin *et al*. introduced a projection gradient method with boundary constraints in [18]. It is simple to calculate and it has better convergence speed than the standard NMF. Zhang *et al*. [21] proposed an rNMF method which decomposes the non-negative data matrix into the sum of the product of a sparse error matrix and two non-negative matrices. In addition, Saakhutdinov *et al*. [22] also improved the convergence of standard NMF.

2.2 Standard NMF Based on Euclidean Distance

The traditional gradient descent method and additive iterative rules were used in the early NMF decomposition, and the negative value must be forced to 0. In [16], Lee and Sueng also proposed to take the square of the Euclidean distance between *V* and *WH* as the objective function, as shown in (8).

$$F = \sum_{i,j} (V_{ij} - (WH)_{ij})^2 \tag{8}$$

The iteration rules are shown in (9) and (10).

$$H_{kj} \leftarrow H_{kj} \frac{(W^T V)_{kj}}{(W^T WH)_{kj}} \tag{9}$$

$$W_{ik} \leftarrow W_{ik} \frac{(VH^T)_{ik}}{(WHH^T)_{ik}} \tag{10}$$

The initial value of *r*, the matrix *W* and matrix *H* should be set properly before iteration. As *V* and *WH* tend to be fixed, the objective function converges. The Euclidean distance based NMF is more popular than the *KL* divergence method. It also reconciled the contradiction between algorithm efficiency and implementation difficulty. In the algorithm given by Lee [16], the initial value of the matrix *W* and *H* can be arbitrary. In addition, three ways of initializing *W* and *H* are mentioned in [15]. The comparison of two standard NMF algorithms are shown in Table 1.

Table 1. The comparison of two standard NMF algorithms

Algorithm	Difficulty	Main characteristics	Multiplicative complexity	Additive complexity
Euclidean Distance NMF	Easy	The contradiction between algorithm efficiency and implementation simplicity is reconciled	6mnr + 2r(m + n)	6mnr
KL Divergence NMF	Easy	The contradiction between algorithm efficiency and implementation simplicity is reconciled	4mnr + 2mn + r(2m + n)	4mnr–2mn–rn

Lee and Seung did some experiments based on this method and the results are shown in Fig. 2. When the column vector of the input matrix is an image of the human face, the basis of NMF shows the local features of facial elements, such as mouth, nose and eyes.

Fig. 2. The image of W in CBCL database obtained by standard NMF

2.3 NMF Algorithms Based on Other Objective Functions

Since Lee proposed the objective function based on the *KL* divergence [2], many researchers have proposed various modified objective functions. Banerjee *et al*. [23] introduced Bregman divergence, which is a more general representation of Euclidean distance and *KL* divergence. Cichocki *et al*. proposed an objective function based on Csiszar's divergence [24]. The flexible and relaxed form of the NMF algorithms were given to increase convergence speed and impose some desired constraints, such as sparsity and smoothness of components. Similarly, Park proposed Young's divergence [25] based on Fenchel inequality, Sun and Fevotte [26] studied the basic theory of beta-divergence. Ding *et al*. [27] proposed that orthogonal nonnegative matrix *t*-factorizations can be obtained by directly adding orthogonal constraints to the objective function. Experimental results showed that this algorithm is comparable to *K*-means algorithm in clustering tasks.

3 Improved NMF Methods

In addition to modifying Lee's standard NMF method with different objective functions, other researchers have focused on extending the NMF function by adding auxiliary constraints to matrix W and H, resulting in a large number of improved NMF algorithms. In this section, the improved NMF algorithms are classified into four categories, which are sparsity enhanced NMF algorithm, discriminative NMF algorithm, manifold based NMF and two-dimensional NMF. Among them, sparsity enhanced NMF algorithms are the most widely used. We will focus on the analysis and comparison of this kind of algorithm.

3.1 Sparsity Enhanced NMF Algorithms

When the subspace method is used for feature extraction, the basis matrix is expected to be as sparse as possible in order to obtain local features with sufficient details. When the coefficient matrix is sparse enough, the image to be measured is only related to a few base image. In this case, features are much easier to be detected. Standard NMF algorithms can be used to describe local features, but sometimes the sparsity of description is not satisfactory. Three typical sparsity enhanced NMF algorithms will be discussed in detail in this section, which are local non-negative matrix factorization (LNMF), NMF algorithm based on gradient descent sparse constraint (NNSC) and sparse non-negative matrix factorization (SNMF).

3.1.1 LNMF

In view of the low accuracy of traditional NMF, as well as the special requirements on the input image in some cases, Li *et al*. [28] proposed a local non-negative matrix factorization (LNMF) algorithm, in which the local feature analysis [29-30] is considered in the objective function. Specifically, three limitations are considered in LNMF. First, try to make the base matrix as sparse as possible. Second, try to minimize the redundancy between different bases. Furthermore, keep only the base that contains the most important information. The object function of LNMF is given in (11):

$$F = \sum_{i,j}(V_{ij}\log\frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij}) - a\sum_{i,j} p_{ij} - b\sum_{i} q_{ii} \tag{11}$$

Where a and b are both constants, $P = (p_{i,j}) = W^T W$, $Q = (q_{i,j}) = HH^T$.

The iteration is carried out according to the following rules.

$$H_{kj} \leftarrow \sqrt{H_{kj}\sum_{i} V_{ij}\frac{W_{ik}}{\sum_k W_{ik}H_{kj}}} \tag{12}$$

$$W_{ik} \leftarrow \frac{W_{ik}\sum_j V_{ij}\frac{H_{kj}}{\sum_k W_{ik}H_{kj}}}{\sum_j H_{kj}} \tag{13}$$

$$W_{ik} \leftarrow \frac{W_{ik}}{\sum_i W_{ik}} \tag{14}$$

Column orthogonal constraint on the basis matrix W is introduced in LNMF. The result is that the sparsity of the basis matrix increased but the sparsity of the coefficient matrix decreased. Experimental results based on face reconstruction and recognition showed that the feature images decomposed by LNMF are sparser than that of NMF, and are superior in terms of recognition rate and training speed [31]. Fig. 3 showed the image of W obtained after LNMF decomposition of samples in the ORL database. Compared with the local characteristics showed in Fig.2, the images after LNMF represented global characteristics of human faces. As a method based on spatial localization, LNMF can be applied to deal with some special cases, such as when an image is partially occluded [32]. LNMF provides an effective way to improve computational efficiency and to reduce the storage requirements in the area of image detection and recognition.

Fig. 3. The image of W in ORL database obtained by LNMF

3.1.2 NNSC

NMF and LNMF have been proved to be effective in the analysis of multivariate data. One of the most important properties is that the decomposition result is a sparse matrix. However, the sparseness of NMF and LNMF is not good enough for the application of NMF in some critical areas. In this case, how to reasonably control the sparseness is of great importance for NMF methods. Hoyer [33] proposed an improved NMF algorithm based on gradient descent sparse constraints (NNSC) [34]. NNSC is the combination of standard NMF and sparse coding. The objective function of NNSC can be described in (15).

$$F = \frac{1}{2}\sum_{i,j}(V_{ij} - (WH)_{ij})^2 + u\sum_{k,j} H_{kj} \tag{15}$$

Where $W_{ik} \geq 0$, $H_{kj} \geq 0$, $\forall\, k, \sum_i W_{ik} = 1$, $u \geq 0$.

The iteration is carried out according to the following rules.

$$W_{ik} \leftarrow W_{ik} - u\sum\nolimits_j ((WH)_{ij} - V_{ij})H_{kj} \tag{16}$$

$$W_{ik} \leftarrow 0, \quad W_{ik} < 0 \tag{17}$$

$$W_{ik} \leftarrow \frac{W_{ik}}{\sqrt{\sum_i W_{ik}^2}} \tag{18}$$

$$H_{kj} \leftarrow H_{kj} \sum\nolimits_i W_{ik} \frac{V_{ij}}{(W_{ik}(WH)_{ij} + u)} \tag{19}$$

Where u is the learning rate. One limition of this algorithm is that the learning of the basis vector is an additive iteration operation in which the non-negativity cannot be guaranteed. Thus the negative value must be forced to be zero.

In application, the form of the second term of the objective function F in (15) can be modified according to certain requirements, such as sparseness, smoothness, etc. It must be carefully selected for the implementation of sparse coding. It is pointed out that when NMF is constrained to obtain the desired sparseness, the first question to be answered is which one should be sparsely restricted, the base vector or the coefficient matrix? The answer to this question depends on the specific problem. It must be determined by appropriate experiments. Therefore, Hoyer proposed to control the L_1 and L_2 measures of the base matrix or coefficient matrix simultaneously for the desired sparseness [33]. The sparsity of both W and H can be adjusted in this method.

3.1.3 SNMF

Based on the fact that the partially-based representation cannot be obtained by NNSC in some complex cases [35-36], NMF is further combined with sparse coding to form sparse non-negative matrix factorization (SNMF) which is especially for better sparse representation. The objective function of SNMF is given in (20).

$$F = \sum_{i,j}(V_{ij}\log(\frac{V_{ij}}{(WH)_{ij}}) - V_{ij} + (WH)_{ij}) + a\sum_{k,j} h_{kj} \tag{20}$$

Where $B, H \geq 0$, $\sum_i W_{ik} = 1$. a is a positive constant for the sparseness constraint of H.

The iteration is carried out according to the following rules.

$$H_{kj} \leftarrow \frac{H_{kj}\sum_i V_{ij}\frac{W_{ik}}{\sum_l W_{il}H_{lj}}}{1+a} \quad \textbf{(21)}$$

$$W_{ik} \leftarrow \frac{W_{ik}\sum_j V_{ij}\frac{H_{kj}}{\sum_l W_{il}H_{lj}}}{\sum_j H_{kj}} \quad \textbf{(22)}$$

$$W_{ik} \leftarrow \frac{W_{ik}}{\sum_i W_{ik}} \quad \textbf{(23)}$$

In addition to ensuring that the decomposition results are non-negative, SNMF also considers the sparseness of the coefficient matrix *H*, so that the original matrix is correlated with only a few basis vectors. There is only multiplication in the iteration of SNMF. This operation keeps the non-negative characteristics of data well, but it also greatly increases the amount of computation. The comparison of LNMF, NNSC and SNMF is shown in Table 2.

Table 2. The comparison of three typical improved NMF algorithms

Algorithm	Sparse control mode	Main characteristics	Difficulty
LNMF	Column orthogonally constraint on the base matrix	The sparsity of the basis matrix increased but the sparsity of the coefficient matrix decreased	Easy
NNSC	Sparse constraints on the coefficient matrix	An effective attempt to combine sparse coding with NMF, insufficient expression of “part”	Easy
SNMF	Sparse constraints on the coefficient matrix	Sparser than NMF with as good expression of “part”	Easy

3.1.4 Other Sparsity Enhanced NMF Algorithms

In order to enhance the sparsity of the base matrix and coefficient matrix in NMF algorithms, many researchers have proposed different methods. In [19], Hoyer uses sparse constraints on *W* and *H* to improve the localized representation of data. This algorithm can control the sparsity of the decomposed low-dimensional matrix. It takes Euclidean distance as the objective function and achieve the precise control of sparsity through nonlinear projection. However, it should be pointed out that when *W* and *H* are subject to high sparsity restrictions, it always has poor description ability on data [37]. Qian *et al*. [38] designed a sparse non-negative matrix algorithm based on $L_{1/2}$ norm constraint to reduce the impact of data noise. Gillis *et al*. [39] discussed the application of Fisher linear discriminant analysis in the improvement of spatial localization characteristics. Guillamet *et al*. [40] eliminated the eigenredundancy of columns in the basis matrix by multiplying the standard NMF and a diagonal matrix that satisfies $VQ = WHQ$, although this problem can also be solved by placing a random constraint on the columns. Additional restrictions on NMF are generally determined by practical problems. Constraints are always applied to compensate for the degradation of data *V* due to noise or other reasons. Recently, Fang *et al*. [41] proposed a deep NMF algorithm by combining deep non-negative matrix factorization with sparsity constraint. It has been well applied in the area of hyperspectral decomposition. Pauca *et al*. [42] added smoothness limitation to deal with the spectral characteristics of remote sensing data. Rutkowski and Cichocki [43] used temporal smoothness and spatial constraints to improve the analysis effect of EEG data.

3.2 Discriminative NMF Algorithms

The unsupervised training of NMF making it unable to make full use of the differential information in the training set. In order to get better classification results, Wang *et al*. [44] proposed Fisher NMF algorithm by adding Fisher discrimination information [45] (the difference between divergence within class and divergence between classes) as penalty term to the objective function. Xue *et al*. [46] proposed FNMF in which Fisher linear discriminant analysis is used to maximize the inter-class dispersion and minimize the intra-class dispersion of coefficient matrix H. FNMF has also been proved to have better performance than PCA, NMF and LNMF in face recognition. Similarly, Bittorf *et al*. [47] proposed an improved NMF algorithm based on linear programming [48], which has good performance on large data sets. Babaee *et al*. [49] proposed the discriminative orthogonal NMF algorithm, which preserves both local and global information of data and achieves good performance in data representation. Dorffer *et al*. [50] proposed nonnegative matrix factorization with Nesterov iterations (Ne-NMF) for better convergence rate.

When the input matrix is not ideal, Li and Wang [51] proposed a multi-layer NMF algorithm (MNMF) and applied it to blind source separation. This is essentially a gradient transformation method that aims to find a solution closer to the global optimum. The main idea is to carry out a series of non-negative matrix factorizations, that is, to decompose the matrix V by NMF, $V = A_1S_1$, and then to decompose the matrix S_1 by NMF, $S_1 = A_2S_2$. So that the NMF can be repeated many times according to the actual problems until the requirements of the objective function are met. This process can be expressed as: $V = A_1 A_2 \ldots A_l S_l$, base matrix $W = A_1 A_2 \ldots A_l$, weight matrix $H = S_l$. Besides, many other works have also started the research of improved NMF in blind source separation [52-58].

3.3 NMF Algorithms Based on Manifold

In order to reduce the dimension of nonlinear data, many NMF algorithms based on manifold features have also been proposed. Cai *et al*. [59] first applied graph regularization method to NMF and proposed graph regularized non-negative matrix factorization (GNMF). In this method, the local neighborhood structure of the original data is used to calculate the adjacency graph matrix, and then the adjacency graph matrix is added to the objective function as a constraint, so that the NMF algorithm can maintain the potential structure of the original data. Subsequently, manifold learning methods are combined with NMF algorithm, such as GDNMF [60], LCGNMF [61] and NLMF [62]. These methods improved the performance of image clustering or classification by constructing the graph Laplacian regularization term. Shang *et al*. [63] proposed dual graph regularization NMF by considering both data manifold and characteristic manifold. Meng *et al*. [64] added sparsity constraints to the graph dual model, which effectively simplified the high-dimensional computation and revealed the bimodal structure between data and features.

It is worth noting that there is not a unique global optimal solution for the NMF problem. Assume that the solution of NMF is W^* and H^*, then W^*D and $D^{-1}H^*$ must also be a solution according to the matrix theory. D and D^{-1} are also non-negative matrices. In this way, there will be an infinite number of solutions. To reduce multiple solutions, some algorithms normalize the columns or rows after each iteration. Monte Carlo method [65] can be used to implement multiple experiments with different initial values, and finally a average result can be obtained.

3.4 Two-dimensional NMF Algorithms

The NMF algorithms mentioned above directly converts two-dimensional (2D) matrix into one dimensional vector. Generally speaking, this transformation is likely to lose the structural information hidden in the original 2D image. To solve these problems, two-dimensional non-negative matrix decomposition (2DNMF) provided an effective way to apply NMF directly on the 2D image matrix. Zhang *et al*. [66] first applied NMF to column vectors and row vectors to obtain the corresponding one-dimensional column vector basis and row vector basis. Then, the cross product of these column vectors and row vectors is calculated as the 2D basis of 2DNMF. The projection coefficient of the image to be measured under the 2D basis matrix is obtained as the image feature. Since 2DNMF can be regarded as the NMF based on row decomposition, it only retains the relevant information between the row vectors of the image, but not all the structural information. By combining the advantages of 2D discrete wavelet transform and 2D non-negative matrix decomposition, Li *et al*. [67] proposed a new face recognition fusion algorithm. The efficiency and recognition rate of this algorithm has been effectively improved. Ftoutou *et al.* [68] studied the supervised classification of diesel engine faults based on 2DNMF and achieved good results. Yang *et al.* [69] proposed two improved 2DNMF algorithms and applied them to face recognition from the perspective of improving the recognition rate and reducing computation.

Experimental results showed that it has better description ability and better recognition effect for both conventional face recognition and occlusion face recognition.

4 Discussion and Conclusion

As a feature analysis method, NMF has been widely used in the area of face recognition, text clustering and image fusion, however, there are still many problems that need further study.

First, under what conditions can NMF get the correct decomposition results? That is, whether the model obtained by NMF training can meet the requirements of hidden structure in the dataset. This is mainly up to the dimension of the basis vector W, i.e. the value of r. In the existing methods, r is basically determined by repeated experiments. How to find the r that satisfies the internal structure of data still needs further research. Second, NMF is sensitive to initial conditions. Generally speaking, the basis vectors obtained by different initial conditions are also different. How do different initial conditions affect the effectiveness and convergence of the algorithm is still not clear. Although many improvements have been proposed to solve this problem, it is still worth further study. Third, the multiplicative iteration method in NMF still has limitations when the size of the dataset is large, we should find a faster convergence iteration method instead. In addition, how to compare the convergence of various NMF methods and how to prove the convergence of these methods still need further improvements. Forth, NMF provides a novel method of feature analysis by matrix decomposition. From the mathematical point of view, can NMF be used for matrices with a large conditional numbers? The matrix with large condition number is often encountered in dealing with practical problems. At present, few references give a clear answer to this question. Moreover, the combination of NMF and other methods also worth exploring. From the physiological and psychological point of view, NMF proves the possibility of understanding the whole through parts, but whether this process is consistent with how human brains understand the world?

Object features can be color, texture, shape or spatial relationship. As a feature analysis method, NMF showed outstanding performance in the reduction of the raw data dimension. The low dimensional features hidden in the internal structure of the data can also be found by NMF. Scientists have done plenty of research on NMF and have put forward many improved methods. The further improvement and application of NMF still relies on the development of new methods and techniques.

Acknowledgements

This work was supported by Beijing Municipal Science Foundation (No. 3214061), General Project of Scientific Research of Beijing Municipal Education Commission (No. KM201911232017) and Qin Xin Talents Cultivation Program, Beijing Information Science & Technology University (QXTCP C202106).

References

[1] K.-J. Wang, C.-T. Zuo, Improvements of non-negative matrix factorization for image detection, Application Research of Computers 31(4)(2014)970-975.

[2] D.-D. Lee, H.-S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401(1999)788-791.

[3] Y.-H. Kuan, C.-M. Kuo, N.-C. Yang, Color-based image salient region segmentation using novel region merging strategy, IEEE Transactions on Multimedia 10(5)(2008)832-845.

[4] S.-M. Chavda, G.-M. Mahesh, Hybrid approach to content-based image retrieval using modified multi-scale LBQ and color features, SN Computer Science 1(6)(2020)1-15.

[5] N.-H. Kim, A.-C. Bovik, S.-J. Aggarwal, Shape description of biological objects via stereo light microscopy, IEEE Transactions on Systems, Man, and Cybernetics 20(2)(2015)475-489.

[6] H. Noh, A. Araujo, J. Sim, T. Weyand, B. Han, Large-scale image retrieval with attentive deep local features, in Proc. 2017 IEEE International Conference on Computer Vision, 2017.

[7] Z. He, F. Nan, X. Li, S.-J. Lee, Y. Yang, Traffic sign recognition by combining global and local features based on semi-supervised classification, IET Intelligent Transport Systems 14(5)(2020)323-330.

[8] J.-J. Koenderink, Theory of edge-detection, Springer Press, Berlin Heidelberg, 2012.

[9] Y. Zhang, W. Zheng, K. Leng, H. Li, Background subtraction using an adaptive local median texture feature in illumination changes urban traffic scenes, IEEE Access (99)(2020)1-10.

[10] K. Kaplan, Y. Kaya, M. Kuncan, M.-R. Mnaz, H. -M. Ertun, An improved feature extraction method using texture analysis with LBP for bearing fault diagnosis, Applied Soft Computing (87)(2020)109-116.

[11] Y. Chen, B. Song, X. Du, M. Guizani, Infrared small target detection through multiple feature analysis based on visual saliency, IEEE Access 7(2019)38996-39004.

[12] P. Paatero, U. Tapper, Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data value. Environmentrics (5)(1994)111-126.

[13] P. Paatero, Least-squares formulation of robust non-negative factor analysis, Chemometrics and Intell Laboratory System (37)(1997)23-35.

[14] D.-D. Lee, H.-S. Seung, Unsupervised learning by convex and conic coding, in Proc. 9th International Conference on Neural Information Processing Systems, 1997.

[15] D. Donoho, Stodden V. When does non-negative matrix factorization give a correct decomposition into parts? Elsevier, 15(2004)1141-1148.

[16] D.-D. Lee, H.-S. Seung, Algorithms for non-negative matrix factorization, in Proc. International Conference on Neural Information Processing Systems, 2001.

[17] E. Gonzalez, Y. Zhang, Accelerating the Lee-Seung algorithm for nonnegative matrix factorization, Technical Report TR-05-02, Rice University, 2005.

[18] C. -J. Lin, Projected gradient methods for non-negative matrix factorization, Technical Report ISSTECH-95-013, National Taiwan University, 2005.

[19] P.-O. Hoyer, Non-negative matrix factorization with sparseness constraints. Journal of Machine Learning Research 5(2004)1457-1469.

[20] P. Paatero, The multilinear engine-a table-driven least squares program for solving multi-linear problems, including the n-way parallel factor analysis model, Computer Graphical Statistics 8(4)(1999)1-35.

[21] L. Zhang, Z. Chen, M. Zheng, X.-F. He, rNMF: robust nonnegative matrix factorization, Frontiers of Electrical & Electronic Engineering in China 6(2)(2015)192-200.

[22] I. Panageas, S. Skoulakis, A. Varvitsiotis, Convergence to second-order stationarity for non-negative matrix factorization: provably and concurrently, (2020) arXiv:2002.11323v2 [cs.LG].

[23] A. Banerjee, S. Merugu, I.-S. Dhillon, J. Ghosh, J. Lafferty, Clustering with bregman divergences, Journal of Machine Learning Research 6(4)(2005)1705-1749.

[24] A. Cichocki, Csiszar's divergences for non-negative matrix factorization: Family of new algorithms, in Proc. 6th International Conference on Independent Component Analysis and Blind Signal Separation, 2006.

[25] M. -Y. Park, T. Hastie, Generalized linear models, Springer Texts in Statistics 69(4)(2007)659-677.

[26] D.-L. Sun, C. Fevotte, Alternating direction method of multipliers for non-negative matrix factorization with the beta-divergence, in Proc. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing, 2014.

[27] C. Ding, T. Li, W. Peng, H. Park, Orthogonal non-negative matrix t-factorizations for clustering, in Proc. 12th ACM International Conference on Knowledge Discovery and Data Mining, 2006.

[28] S.-Z Li, W.-H Xin, H.-J Zhang, Q.-S Cheng. Learning spatially localized, parts-based representation, in Proc. IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2001.

[29] Y. Meng, R. Shang, L. Jiao, W. Zhang, Y. Yuan, S. Yang, Feature selection based dual-graph sparse non-negative matrix factorization for local discriminative clustering, Neuro computing 290(17)(2018)87-99.

[30] D. -A. Abdullah, M. -H. Akpnar, A. Engür, Local feature descriptors based ECG beat classification, Health Information Science and Systems 8(1)(2020)1-14.

[31] L. Li, Y.-J. Zhang. A survey on algorithms of non-negative matrix factorization, Acta Electronic Sinica 36(004) (2008) 737-743.

[32] R. Shang, J. Song, L. Jiao, Double feature selection algorithm based on low-rank sparse non-negative matrix factorization, International Journal of Machine Learning and Cybernetics 11(2)(2020)1-18.

[33] P. -O. Hoyer, Non-negative sparse coding, in Proc. 12th IEEE Workshop on Neural Networks for Signal Processing, 2002.

[34] J. -J. Liu, Research on sparse code shrinkage denoise in underwater 3d laser scanning images. in Proc. 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference, 2019.

[35] Y. Peng, Y. Zhang, F. Qin, W. Kong, Joint non-negative and fuzzy coding with graph regularization for efficient data clustering, Egyptian Informatics Journal 22(1)(2021)91-100.

[36] X. Li, Y. Guan, Y. Wu, Z. Zhang, Piano multi-pitch estimation using sparse coding embedded deep learning, EURASIP Journal on Audio Speech and Music Processing 1(2018)132.

[37] X. Chen, L. Gu, S.-Z Li, H.-J. Zhang. Learning representative local features for face detection, in Proc. IEEE Computer Society Conference on Computer Vision & Pattern Recognition. 2001.

[38] Y. Qian, S. Jia, J. Zhou, A. Robleskelly, Hyperspectral unmixing via sparsity-constrained nonnegative matrix factorization, IEEE Transactions on Geoscience & Remote Sensing 49(11)(2011)4282-4297.

[39] N. Gillis, Introduction to non-negative matrix factorization, SIAG/OPT Views and News 25(1)(2017)7-16.

[40] D. Guillamet, M. Bressan, J. Vitria, A weighted non-negative matrix factorization for local representations, Computer Vision and Pattern Recognition 1(2001)942-947.

[41] H. Fang, A. Li, H. Xu, T. Wang, Sparsity-constrained deep nonnegative matrix factorization for hyperspectral unmixing, IEEE Geoence & Remote Sensing Letters 5(2018) 1105-1109.

[42] P. Pauca, J. Piper, R. Plemmons, Nonnegative matrix factorization for spectral data analysis. Linear Algebra Applications 416(1)(2006) 29-47.

[43] T. -M. Rutkowski, R. Zdunek, A. Cichocki, Multichannel EEG brain activity pattern analysis in time-frequency domain with nonnegative matrix factorization support. International Congress Series 13(1)(2007)266-269.

[44] G. Wang, A.-V. Kossenkov, M.-F Ochs. LS-NMF: A modified non-negative matrix factorization algorithm utilizing uncertainty estimates, Bmc Bioinformatics 7(2006)344-349.

[45] Z. Guo, A. Min, B. Yang, J.-H. Chen, H. Li, A modified Huber nonnegative matrix factorization algorithm for hyperspectral unmixing, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 99(2021)1-6.

[46] X. Bai, C. Wang, Color face recognition based on NMF with Fisher rule, in Proc. 2nd IEEE International Conference on Information Management and Engineering, 2010.

[47] V. Bittorf, B. Recht, C. Re, J.-A Tropp, Factoring nonnegative matrices with linear programs, Advances in Neural Information Processing Systems 2(2012)1214-1222.

[48] B. Cantó, R. Cantó, A.-M. Urbano, Irreducible totally nonnegative matrices with a prescribed Jordan structure, Linear Algebra and its Applications 609(2021)129-151.

[49] M. Babaee, S. Tsoukalas, M. Babaee, G.Rigoll, M. Datcu, Discriminative nonnegative matrix factorization for dimensionality reduction, Neurocomputing 173(2015) 212-223.

[50] C. Dorffer, M. Puigt, G. Delmaire, G. Roussel, Fast nonnegative matrix factorization and completion using Nesterov iterations, Springer International Publishing, 2017.

[51] F. Li, W. Xu, Hyperspectral unmixing using double-constrained multilayer NMF, Remote Sensing Letters 10(3)(2019)224-233.

[52] A. Mirzal. NMF versus ICA for blind source separation, Advances in Data Analysis and Classification 11(2017)25-48.

[53] A. -N. H, B. D. A. R, A comparative study of blind source separation for bioacoustics sounds based on fast ICA, PCA and NMF- science direct, Procedia Computer Science, 126 (2018) 363-372.

[54] A. Rolet, V. Seguy, M. Blondel, H. Sawada, Blind source separation with optimal transport non-negative matrix factorization, Journal on Advances in Signal Processing 7(2018)53-60.

[55] A. Pg, B. Sv, Upgrading sparse NMF algorithm for blind source separation through adaptive parameterized hybrid kernel based approach, Measurement 143(2019)11-21.

[56] H. Hu, G. Li, M. -E. Lutman, S. Bleeck, Nonnegative matrix factorization sparse coding strategy for cochlear implants, Springer Press, Berlin Heidelberg, 2014.

[57] K. -S. Santosh, S. -H. Bharathi, Non-negative matrix factorization algorithms for blind source separation in speech recognition, in Proc. 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology, 2017.

[58] V. Leplat, N. Gillis, A. Ang, Blind audio source separation with minimum-volume beta-divergence NMF, IEEE Transactions on Signal Processing 68(2020)3400-3410.

[59] D. Cai, X. He, J. Han, T. -S. Huang, Graph regularized nonnegative matrix factorization for data representation, IEEE Transactions on pattern analysis and machine intelligence 33(8)(2011)1548-1560.

[60] X. Long, H. Lu, Y. Peng, W. Li, Graph regularized discriminative non-negative matrix factorization for face recognition, Multimedia tools and applications 72(3)(2014)2679-2699.

[61] Q. Liao, Q. Zhang, Local coordinate based graph-regularized NMF for image representation, Signal processing 124(9)(2016)103-114.

[62] X. Li, G. Cui, Y. Dong, Graph regularized non-negative low-rank matrix factorization for image clustering, IEEE transactions on cybernetics 47(11)(2017)3840-3853.

[63] F. Shang, L. -C. Jiao, F. Wang, Graph dual regularization non-negative matrix factorization for co-clustering, Pattern recognition 45(6)(2012)2237-2250.

[64] Y. Meng, R. Shang, L. Jiao, Y. Yuan, S. Yang, Feature selection based dual-graph sparse non-negative matrix factorization for local discriminative clustering, Neurocomputing 290(23)(2018)87-99.

[65] R.-M. Ger Be, N. Chopin, Sequential Quasi-Monte Carlo, Journal of the Royal Statistical Society 77(3)(2015)509-579.

[66] D. Zhang, S. Chen, Z.-H. Zhou, Two-dimensional non-negative matrix factorization for face representation and recognition, in Proc. 2nd international workshop on analysis and modeling of faces and gestures, 2005.

[67] K.-Z. Li, B.-H. Wang, H. Lou, Y. Zheng, Face recognition algorithm based on wavelet transform and two-dimensional non-negative matrix factorization, Application Research of Computers 30(004)(2013)1275-1277.

[68] E. Ftoutou, M. Chouchane, Feature extraction using s-transform and 2dNMF for diesel engine faults classification, Springer International Publishing, 2017.

[69] J.-J. Yang, Research on face recognition algorithm based on two-dimensional NMF, Dalian Maritime University, 2016.