



Max–min distance nonnegative matrix factorization



Jim Jing-Yan Wang, Xin Gao*

Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

HIGHLIGHTS

- A novel supervised nonnegative matrix factorization method is proposed.
- Within-class and between-class pairs are defined by class labels.
- The maximum within-class distance is minimized in NMF space.
- The minimum between-class distance is maximized in NMF space.
- Experiment results show its outperformance over other supervised NMF methods.

ARTICLE INFO

Article history:

Received 10 February 2014
Received in revised form 12 October 2014
Accepted 16 October 2014
Available online 26 October 2014

Keywords:

Data representation
Nonnegative matrix factorization
Supervised learning
Max–min distance analysis

ABSTRACT

Nonnegative Matrix Factorization (NMF) has been a popular representation method for pattern classification problems. It tries to decompose a nonnegative matrix of data samples as the product of a nonnegative basis matrix and a nonnegative coefficient matrix. The columns of the coefficient matrix can be used as new representations of these data samples. However, traditional NMF methods ignore class labels of the data samples. In this paper, we propose a novel supervised NMF algorithm to improve the discriminative ability of the new representation by using the class labels. Using the class labels, we separate all the data sample pairs into within-class pairs and between-class pairs. To improve the discriminative ability of the new NMF representations, we propose to minimize the maximum distance of the within-class pairs in the new NMF space, and meanwhile to maximize the minimum distance of the between-class pairs. With this criterion, we construct an objective function and optimize it with regard to basis and coefficient matrices, and slack variables alternatively, resulting in an iterative algorithm. The proposed algorithm is evaluated on three pattern classification problems and experiment results show that it outperforms the state-of-the-art supervised NMF methods.

© 2014 The Authors. Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

1. Introduction

Nonnegative matrix factorization (NMF) has attracted much attention from both research and engineering communities (Eches & Guillaume, 2014; Lin, 2007; Malley, Braban, & Heal, 2014; Seung & Lee, 2001; Vidar & Alvindia, 2013; Wang, Almasri, & Gao, 2012; Wang, Bensmail, & Gao, 2013; Wang & Gao, 2014; Zheng, Zhang, Ng, Shiu, & Huang, 2011). Given a data matrix whose elements are all nonnegative, NMF tries to decompose it as the product of two nonnegative low-rank matrices. One matrix can be regarded as a basis matrix with its columns as basis vectors, and the other one as a linear combination coefficient matrix, so that the original

data columns in the original matrix could be represented as the linear combination of the basis vectors. Because of the nonnegative constraints on both the factorization metrics, it only allows the additive linear combination, and thus a part-based representation could be achieved (Agarwal, Awan, & Roth, 2004; Cai, He, Han, & Huang, 2011; Hwang & Kang, 2013; Lemme, Reinhart, & Steil, 2012; Zhao, Li, Wu, Fu, & Liu, 2013). Since the original NMF approach was proposed by Lee and Seung (1999) and Seung and Lee (2001), due to its ability to learn the parts of the data set (Li, Hou, Zhang, & Cheng, 2001), it has been used as an effective data representation method in various problems, such as pattern recognition (Hoyer, 2004; Liu, Zheng, & You, 2006; Van Hamme, 2012; Zhu, 2008), computer vision (Guillamet, Vitri, & Schiele, 2003; Monga & Mihak, 2007; Shashua & Hazan, 2005), and bioinformatics (Gao & Church, 2005; Pascual-Montano, 2008; Tian, Liu, & Wu, 2013). The most popular application of NMF as a data representation tool is in pattern recognition, where the nonnegative feature vectors of

* All correspondence should be addressed to Xin Gao. Tel.: +966 12 8080323.

E-mail addresses: jimjywang@gmail.com (J.J.-Y. Wang), xin.gao@kaust.edu.sa (X. Gao).

the data samples are organized as a nonnegative matrix, and the columns of the coefficient matrix are used as the new low-dimensional representations.

In the pattern recognition problems, when NMF is applied on the data matrix, it is usually assumed that the class labels of the data samples are not available, making it an unsupervised problem (Mohammadiha, Smaragdis, & Leijon, 2013; Tsarev, Petrovskiy, & Mashechkin, 2011). Some typical applications include clustering of images and documents (Cai et al., 2011; Liu, Wu, Li, Cai, & Huang, 2012). However, in real world supervised or semi-supervised classification applications, class labels of training data samples are usually available, which is ignored by most existing NMF methods. If the class label information could be utilized during the representation procedure, the discriminative ability of the representation could be significantly improved (Gaujoux & Seoighe, 2012; Kitamura et al., 2013; Zhang, Xia, Yang, & Yang, 2007; Zhou & Schwenker, 2013). To this end, some supervised and semi-supervised NMF methods were proposed. For example, Wang and Jia (2004) proposed the Fisher nonnegative matrix factorization (FNMF) method to encode discrimination information for a classification problem by imposing Fisher constraints on the NMF algorithm. Lee, Yoo, and Choi (2010) proposed the semi-supervised nonnegative matrix factorization (SSNMF) by jointly incorporating the data matrix and the partial class label matrix into NMF. Most recently, Liu, Wu et al. (2012) proposed the constrained nonnegative matrix factorization (CNMF) by incorporating the label information as additional constraints.

In this paper, we propose a novel supervised NMF method, by exploring the class label information and using it to constrain the learning of coefficient vectors of the data samples. We consider pairs of data samples, and the class labels of the samples allow us to separate the pairs to two types—the within-class pair and the between-class pair. The within-class pair refers to a pair of samples with the same class label, while the between-class pair refers to a pair of samples with different class labels. To improve the discriminative ability of the coefficient vectors of the samples, we consider the distance between the coefficient vectors of each sample pairs, and try to minimize that of the within-class pairs, while maximize that of the between-class pairs. In this way, the coefficient vectors of data samples of the same class can be gathered, while that of different classes can be separated. One problem is how to assign different weights to different pairs in the objective function. To avoid this problem, we apply a strategy similar to max–min distance analysis (Bian & Tao, 2011). The *maximum* within-class pair coefficient vector distance is minimized, so that all the within-class pair coefficient vector distances can be minimized as well. Meanwhile the *minimum* between-class pair coefficient vector distance is maximized, so that all the between-class pair coefficient vector distances can be maximized as well. We construct a novel objective function for NMF to impose both the maximum within-class pair distance minimization and the minimum between-class pair distance maximization problems. By optimizing it with an alternative strategy, we develop an iterative algorithm. The proposed method is called Max–Min Distance NMF (MMDNMF).

The remaining parts of this paper are organized as follows: in Section 2, we introduce the novel NMF method. In Section 3, the experimental results are given to verify the effectiveness of the proposed method. The paper is concluded in Section 4.

2. Proposed method

In this section, we first formulate the problem with an objective function, and then optimize it to obtain an iterative learning algorithm.

2.1. Problem formulation

Supposing we have n data samples in a training set $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^n$, where $\mathbf{x}_i \in \mathbb{R}_+^d$ is the d -dimensional nonnegative feature vector of the i th sample, we organize the samples as a nonnegative matrix $X = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}_+^{d \times n}$. The i th column of the matrix X is the feature vector of the i th sample. Their corresponding class label set is denoted as $\{y_i\}_{i=1}^n$, where $y_i \in \mathcal{Y}$ is the class label of the i th sample, and $\mathcal{Y}$ is the class label space. NMF aims to find two low rank nonnegative matrices $U \in \mathbb{R}_+^{d \times m}$ and $V \in \mathbb{R}_+^{m \times n}$, where $m \leq d$, so that the product of them, UV , could approximate the original matrix, X , as accurately as possible,

$$X \approx UV. \quad (1)$$

The m columns of the matrix U could be regarded as m basis vectors, and each sample $\mathbf{x}_i$ could be represented as the nonnegative linear combination of these basis vectors. The linear combination coefficient vector of $\mathbf{x}_i$ is the i th column vector $\mathbf{v}_i \in \mathbb{R}_+^m$ of V . We can also regard $\mathbf{v}_i$ as a new low-dimensional presentation vector of $\mathbf{x}_i$ with regard to the basis matrix U . To seek the optimal matrices U and V , we consider the following problems to construct our objective function:

- To reduce the approximation error between X and UV , the squared ℓ_2 distance between them is usually minimized with regard to U and V as follows,

$$\begin{aligned} \min_{U, V} \|X - UV\|_2^2 \\ \text{s.t. } U \geq 0, V \geq 0. \end{aligned} \quad (2)$$

- We consider the training sample pairs in the training set, and separate them to two pair sets—the within-class pair set $\mathcal{W}$ and the between-class pair set $\mathcal{B}$. The within-class pair set is defined as the set of sample pairs belonging to the same class, i.e., $\mathcal{W} = \{(i, j) | y_i = y_j, \mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}\}$. The between-class pair set is defined as the set of sample pairs belonging to different classes, i.e., $\mathcal{B} = \{(i, j) | y_i \neq y_j, \mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}\}$. To compare the two samples of the (i, j) th pair in the new coefficient vector space, we use the squared ℓ_2 norm distance between their coefficient vectors, $\|\mathbf{v}_i - \mathbf{v}_j\|_2^2$. Apparently, to improve the discriminative ability of the new NMF presentation, the coefficient vector distances of within-class pairs should be minimized, while those of the between-class pairs should be maximized. Instead of considering all the pairs, we directly minimize the maximum coefficient vector distance of the within-class pairs, as follows,

$$\begin{aligned} \min_V \left\{ \max_{(i, j) \in \mathcal{W}} \|\mathbf{v}_i - \mathbf{v}_j\|_2^2 \right\} \\ \text{s.t. } V \geq 0, \end{aligned} \quad (3)$$

and thus we duly consider the aggregation of all within-class pairs. Meanwhile, we also maximize the minimum coefficient vector distance of the between-class pairs, as follows,

$$\begin{aligned} \max_V \left\{ \min_{(i, j) \in \mathcal{B}} \|\mathbf{v}_i - \mathbf{v}_j\|_2^2 \right\} \\ \text{s.t. } V \geq 0 \end{aligned} \quad (4)$$

and thus we consider the separation of all between-class pairs. In this way, the maximum within-class pair distance is minimized, so that all the within-class pair distances are also minimized. Similarly, the minimum between-class pair distance is maximized, so that all the between-class pair distances are also maximized.

Download English Version:

<https://daneshyari.com/en/article/6863325>

Download Persian Version:

<https://daneshyari.com/article/6863325>

[Daneshyari.com](https://daneshyari.com)