Engineering Applications of Artificial Intelligence 69 (2018) 24–35

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Dual-graph regularized non-negative matrix factorization with sparse and orthogonal constraints

Yang Meng [a], Ronghua Shang [a,*], Licheng Jiao [a], Wenya Zhang [b], Shuyuan Yang [a]

[a] Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an, Shaanxi Province 710071, China
[b] School of Computer Sciences, Xidian University, Xi'an, Shaanxi Province 710071, China

ARTICLE INFO

Keywords:
Semi-supervised non-negative matrix factorization
Dual-graph model
Orthogonal constraint
Bi-orthogonal constraints
Cluster

ABSTRACT

Semi-supervised Non-negative Matrix Factorization (NMF) can not only utilize a fraction of label information, but also effectively learn local information of the objectives, such as documents and faces. Semi-supervised NMF is an efficient technique for dimensionality reduction of high dimensional data. In this paper, we propose a novel semi-supervised NMF, called Dual-graph regularized Non-negative Matrix Factorization with Sparse and Orthogonal constraints (SODNMF). Dual-graph model is added into semi-supervised NMF, and the manifold structures of the data space and the feature space are taken into account simultaneously. In addition, the sparse constraint is used in SODNMF, which can simplify the calculation and accelerate the processing speed. The most important is that SODNMF makes use of bi-orthogonal constraints, which can avoid the non-correspondence between images and basic vectors. Therefore, it can effectively enhance the discrimination and the exclusivity of clustering, and improve the clustering performance. We give the objective function, the iterative updating rules and the convergence proof. Empirical experiments demonstrate encouraging results of our novel algorithm in comparison to four algorithms within some state-of-the-art algorithms through a set of evaluations based on three real datasets.

1. Introduction

With the advent of big data era, the amount of data increases more and more, and the dimensionality of data becomes larger and larger (Shang et al., 2016a; Lee and Seung, 1999). How to deal with massive high dimensional data and find an appropriate low dimensional representation of data are important issues in data mining and analysis (Shang et al., 2017, 2016b). Such as the applications of data mining and analysis in commodity recommendation and security monitoring (Tian and Chen, 2017). An effective low dimensional representation of data can not only mine the latent structural information of data (Ma et al., 2016), but also remove the redundant features in the original data and rapidly deal with massive high dimensional data (Gu et al., 2017).

Matrix factorization is an efficient dimensionality reduction technique, which can reduce the dimensionality of high dimensional data. There are many classical matrix factorization techniques, such as Singular Value Decomposition (SVD) (Duda et al., 2012), Principal Component Analysis (PCA) (Jolliffe, 1986), Linear Discriminant Analysis (LDA) (Pang et al., 2014), Non-negative Matrix Factorization (NMF) (Zheng et al., 2007; Paatero and Tapper, 1994) and so on. However, compared with other matrix factorization technique, NMF can effectively learn local information of the objectives, such as documents and faces. Therefore, NMF can make better performance on document clustering (Lee and Seung, 1999; Paatero and Tapper, 1994; Xu et al., 2003; Shahnaz et al., 2006), face recognition (Lee and Seung, 1999; Li et al., 2001) and other practical applications.

NMF aims to decompose the original high dimensional data matrix into two low dimensional data matrices, and the product of the two low dimensional data matrices approximates the original high dimensional data matrix as far as possible. In this way, we can reduce the dimensionality of the original high dimensional data. Classical NMF is an unsupervised learning algorithm, which has been widely used in data clustering. However, there is often a fraction of label information in the original data in the real world, and the classical NMF algorithms cannot make full use of the label information in the original data. Many machine learning researchers have found that the semi-supervised algorithm using a fraction of label information can improve the accuracy of learning (Belkin et al., 2006; Feng et al., 2016; Yang et al., 2014), so the accuracy of unsupervised NMF learning is inferior to many

* Corresponding author.
E-mail address: rhshang@mail.xidian.edu.cn (R. Shang).

https://doi.org/10.1016/j.engappai.2017.11.008
Received 9 May 2017; Received in revised form 4 September 2017; Accepted 24 November 2017

semi-supervised algorithms. Recently, in order to make full use of the label information, many scholars have improved the unsupervised NMF and proposed many semi-supervised NMF algorithms (Liu et al., 2012; Babaee et al., 2016), which can not only take advantages of the local information of the object in NMF, but also effectively utilize a fraction of label information to improve the accuracy of NMF learning. In Liu et al. (2012), Liu et al. proposed a semi-supervised NMF called Constrained Non-negative Matrix Factorization (CNMF) which embeds the label information as hard constraints into the objective function of NMF. In the new low dimensional representation space, the points with the same label have the same coordinates. In Discriminative Nonnegative Matrix Factorization (DNMF) (Babaee et al., 2016), a fraction of label information is introduced by coupling discriminative regularizer to the objective function of the semi-supervised NMF. However, many algorithms like CNMF and DNMF cannot exploit the local geometric information of data and make full use of the potential structural information. In addition, the previous semi-supervised NMF algorithms do not take advantages of the sparsity of matrices, which results in the complex calculation and the long optimization time. The non-correspondence between images and basic vectors also makes the previous semi-supervised NMF algorithms lack of discrimination.

In order to solve the above problems, we propose a novel semi-supervised non-negative matrix factorization algorithm, called Dual-graph regularized Non-negative Matrix Factorization with Sparse and Orthogonal constraints (SODNMF). Motivated by recent progress in dual regularization (Sindhwani et al., 2009; Gu and Zhou, 2009) and structural information (Ma et al., 2016; Gu et al., 2017), we combine dual-graph model with semi-supervised NMF to make full use of the potential structural information, so the manifold structures of the data space and the feature space are taken into account simultaneously. In addition, inspired by the recent development of sparse constraint (Luo and Zhang, 2014) and orthogonal constraint (Ding et al., 2006), we introduce sparse constraint and bi-orthogonal constraints into semi-supervised NMF, which can not only overcome the disadvantage of the slow optimization and the complex calculation in many existing semi-supervised NMF algorithms, but also avoid the non-correspondence between images and basic vectors to effectively enhance the discrimination and the exclusivity of clustering. In order to prove the efficiency and effectiveness of our algorithm, we give the convergence proof of the objective function and the experimental results of SODNMF and other related algorithms on three real datasets ORL, PIE and TDT2.

Our main contributions are the following four aspects:

1. Use of the iterative updating rules derived from ordinary non-negative matrix factorization incorporating all the constraints may not induce a reliable solution due to scaling issue in several low dimensional matrices. To avoid this problem, we introduce a diagonal scaling matrix into semi-supervised NMF.
2. Dual-graph model is added into semi-supervised NMF, which constructs two neighbor graphs of the data space and the feature space respectively. In this way, we can make full use of the potential structural information on account of preserving the manifold structures of the data space and the feature space simultaneously.
3. Sparse constraint with $L_{2,1/2}$-norm on the coefficient matrix in the feature space is incorporated as the additional condition, which can not only make the coefficient matrix with a good sparsity and simplify the calculation, but also enhance the local learning ability and robustness of the algorithm.
4. Bi-orthogonal constraints are adopted in SODNMF. Each image can correspond to the unique basic vector with the orthogonal constraint on the coefficient matrix in the feature space, which can effectively enhance the discrimination of clustering. The orthogonal constraint on the basic matrix in the data space can enhance the exclusivity across the classes and improve the clustering performance.

The rest of the paper is organized as follows: In Section 2, we give an introduction of the classical NMF and some related NMF algorithms. In Section 3, we introduce the mathematical model and the solution procedure of our algorithm, and then prove the convergence of the optimization scheme. In Section 4, we provide a large number of experiments to demonstrate the efficiency and effectiveness of our algorithm. Finally, we draw a conclusion and provide suggestions for future work in Section 5.

2. Related work

2.1. NMF

NMF can obtain two low dimensional data matrices by decomposing the original high dimensional data matrix, and the product of the two low dimensional data matrices approximates the original data matrix as far as possible to find an appropriate low dimensional representation of the original data matrix. We have an original data matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n] \in \Re^{m\times n}$, where m is the number of the feature dimensions, n is the number of the samples. $\boldsymbol{x}_i = [x_{i1}, x_{i2}, \dots, x_{im}]^T \in \Re^m$ is the ith vector. In NMF, the original high dimensional data matrix $\boldsymbol{X}$ should be decomposed into two low dimensional data matrices $\boldsymbol{U} = [u_{ij}] \in \Re^{m\times k}$ and $\boldsymbol{V} = [v_{ij}]^T \in \Re^{n\times k}$, where k is the clustering number, $\boldsymbol{U}$ is the coefficient matrix in the feature space and $\boldsymbol{V}$ is the basic matrix in the data space. The purpose of NMF is to let the product of the coefficient matrix $\boldsymbol{U}$ and the basic matrix $\boldsymbol{V}$ approximate the original data matrix $\boldsymbol{X}$ as far as possible:

$$\boldsymbol{X} \approx \boldsymbol{U}\boldsymbol{V}^T. \tag{1}$$

In other words, we should minimize the following residual error matrix:

$$O = \|\boldsymbol{X} - \boldsymbol{U}\boldsymbol{V}^T\|_F^2 \quad s.t.\ u_{ij} \geq 0, v_{ij} \geq 0 \tag{2}$$

where $\|\cdot\|_F$ is Frobenius norm (F-norm) of the matrix. We can get the Euclidean distance of two matrices by calculating the square of the F-norm. In Zheng et al. (2007), Lee et al. provided the iterative updating rules to solve such a minimization problem, and prove the convergence. The iterative updating rules of the coefficient matrix $\boldsymbol{U}$ and the basic matrix $\boldsymbol{V}$ of NMF are as follows:

$$u_{ij} = u_{ij}\frac{(\boldsymbol{X}\boldsymbol{V})_{ij}}{(\boldsymbol{U}\boldsymbol{V}^T\boldsymbol{V})_{ij}} \tag{3}$$

$$v_{ij} = v_{ij}\frac{(\boldsymbol{X}^T\boldsymbol{U})_{ij}}{(\boldsymbol{V}\boldsymbol{U}^T\boldsymbol{U})_{ij}}. \tag{4}$$

First of the iterative process, we initialize the coefficient matrix $\boldsymbol{U}$ and the basic matrix $\boldsymbol{V}$ randomly, and then update them according to the iterative updating rules in formula (3) and (4) until the final condition is reached.

2.2. GNMF

In Cai et al. (2011), Cai et al. proposed Graph Regularized Nonnegative Matrix Factorization (GNMF) which adds manifold learning into the classical NMF (Belkin et al., 2006; Cai et al., 2009a, b). GNMF constructs a neighbor graph to simulate the local geometric structure of data. For n samples, a neighbor graph with n vertices is constructed, and each vertex corresponds to a sample. For the vertex x_i, we aim to find its k-nearest neighbors and establish the edges and weights with x_i which represent the similarities between them. Therefore, the weight matrix $\boldsymbol{W}$ is also called the similarity matrix. There are many methods to construct the weight matrix in the neighbor graph, the common methods are (Cai et al., 2011): 0–1 weighting, heat kernel weighting and dot-product weighting. Then, $Tr(\boldsymbol{V}^T\boldsymbol{L}\boldsymbol{V})$ can be used to measure the smoothness of the low dimensional representation, so the objective function of GNMF is measured as follows:

$$O = \|\boldsymbol{X} - \boldsymbol{U}\boldsymbol{V}^T\|_F^2 + Tr(\boldsymbol{V}^T\boldsymbol{L}\boldsymbol{V}) \quad s.t.\ u_{ij} \geq 0, v_{ij} \geq 0. \tag{5}$$