Neurocomputing 000 (2018) 1–13

Contents lists available at ScienceDirect

Neurocomputing

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

Feature selection based dual-graph sparse non-negative matrix factorization for local discriminative clustering

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

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

ARTICLE INFO

Article history:
Received 20 November 2016
Revised 10 March 2017
Accepted 6 February 2018
Available online xxx

Communicated by Feiping Nie

keywords:
Dual-graph
Non-negative matrix factorization
Local discriminative
Feature selection
Clustering

ABSTRACT

Non-negative matrix factorization (NMF) can map high-dimensional data into a low-dimensional data space. Feature selection can eliminate the redundant and irrelevant features from the alternative features. In this paper, we propose a feature selection based dual-graph sparse non-negative matrix factorization (DSNMF) which can find an appropriate low dimensional representation of data by NMF and then select more discriminative features to further reduce the dimension of the low dimensional space by feature selection rather than reduce the dimension by only NMF or feature selection in many previous methods. DSNMF combines dual-graph model with non-negative matrix factorization, which can not only simultaneously preserve the geometric structures in both the data space and the feature space, but also make the two non-negative matrix factors update iteratively and interactively. In addition, DSNMF exerts $L_{2,1}$-*norm* constraint on the non-negative matrix factor of the feature space to make full use of the sparse self-representation information. What's more, we propose a new local discriminative feature selection clustering called feature selection based dual-graph sparse non-negative matrix factorization for local discriminative clustering (DSNMF-LDC) whose clustering effects are better. We give the objective function, the iterative updating rules and the convergence proof. Our empirical study shows that DSNMF-LDC is robust and excellent in comparison to 9 feature selection algorithms and 7 clustering algorithms in clustering accuracy (ACC) and normalized mutual information (NMI).

1. Introduction

With the rapid development of information technology and computer, the size of the collected data in many fields has increased. The massive high-dimensional data have put forward severe challenge to the traditional machine learning and the statistical analysis [1]. Feature selection aims to eliminate the redundant and irrelevant features, identify and preserve the discriminative features from the high-dimensional data [2–5]. It can reduce the feature dimensions, simplify the calculation model and improve the model accuracy, running efficiency and learning performance [6]. The discriminative features got from feature selection can be used in clustering to improve the quality of clustering. How to explore the inherent rules and the essential structure in the high-dimensional data, to efficiently obtain the useful features and represent them in low dimensions have become hot issues in the machine learning, pattern recognition, data mining and statistical analysis. It has wide applications in people's life, such as text classification [7], medical diagnosis [8,9], video event detection [10], intrusion detection [11] and some other fields.

Many studies about feature selection algorithms have shown that the data information generally distributes in the nonlinear low-dimensional submanifold of the high-dimensional space, so the researchers put forward a lot of manifold learning methods to discover the potential geometric structure [12–17]. The main idea of Laplacian score (LapScore) [12] and spectral feature selection (SPEC) [13] is to evaluate each feature according to the local preserving strategy and remove the poor features. However, they do not use learning mechanism. Multi-cluster feature selection (MCFS) [14] and minimum redundancy spectral feature selection (MRSF) [15] combine the embedded learning with the sparse constraints, and their difference lies in the different sparse constraints. MCFS uses L_1-*norm* constraint and MRSF uses advanced $L_{2,1}$-*norm* constraint. However, MCFS and MRSF both belong to the step-by-step feature selection, which cannot take the effect of the manifold information on the following feature selection into account. Joint embedding learning and sparse regression feature selection (JELSR) [16] combines the embedded learning with spectral regression, which can effectively preserve the discriminative features. Locality and similarity preserving embedding feature

* Corresponding author.
E-mail addresses: rhshang@mail.xidian.edu.cn, 280331025@qq.com (R. Shang).

https://doi.org/10.1016/j.neucom.2018.02.044
0925-2312/

selection (LSPE) [17] combines the embedded learning with feature selection, which can preserve the discriminative features and excavate the geometric information of the data space. The aforementioned methods can only preserve the local manifold information of the data space and cannot make full use of the local manifold information of the feature space, which cannot completely excavate the potential information. In Ref. [18], Chang et al. have proposed a convex sparse principal component analysis (CSPCA) algorithm which adopts the recent advances of sparsity and robust PCA into a joint framework to leverage the mutual benefit. CSPCA is the first convex sparse and robust PCA algorithm, which can always ensure the algorithm achieves the global optimum. In Ref. [1], an advanced self-representation based dual-graph regularized feature selection clustering (DFSC) has been proposed to solve this problem. DFSC utilizes the dual-graph (the data graph and the feature graph) model which considers the manifold information both of the data space and the feature space to make full use of the geometric structure of the data. DFSC obviously outperforms the previous algorithms at the clustering effects. However, from the updating rules in DFSC, we can see the self-representation coefficients matrices in feature space and data space can only update by themselves rather than affect each other for the self-representation model, which cannot give full play to the dual-graph model.

Clustering is divided into several categories according to the preset clustering number, so as to make the similarities of elements in the same class as large as possible, and make the similarities of elements in the different classes as small as possible [19–21]. To evaluate the performance of feature selection algorithms, we often need to cluster according to the selected features. There are some common clustering algorithms, such as K-means, NMF [22,23], dual regularized co-clustering (DRCC) [24], concept factorization (CF) [25], locally consistent concept factorization (LCCF) [26] and dual-graph regularized concept factorization clustering (GCF) [27]. K-means is one of the commonest clustering algorithms, which is simple and easy to understand, but its performance will significantly decrease in dealing with high-dimensional problems. To solve this problem, linear discriminant analysis (LDA) has been combined with K-means, which can effectively use the discriminative information and achieve good results in data clustering. The purpose of NMF [22,23] is to decompose the input data into two low-dimensional matrices of the data space and the feature space. Inspired by NMF, DRCC [24] does tri-factorization on the input data. In addition, DRCC adopts the dual-graph model to effectively utilize the potential information. CF [25] is an extension of NMF and applies the idea of the kernel method [26], which can be used in the datasets containing negative values. Based on CF, Cai et al. have proposed LCCF [27] which can preserve the geometrical manifold structures. Ye and Jin have proposed GCF [28] which adds dual-graph model into LCCF. GCF can preserve the geometrical manifold information both of the data graph and the feature graph. However, the aforementioned clustering algorithms do not apply the feature selection in advance when dealing with high-dimensional data, so there are some redundant and irrelevant features in the original features.

To solve the problem in the aforementioned feature selection algorithms [12–17], we propose a feature selection based dual-graph sparse non-negative matrix factorization (DSNMF). Dual-graph model is added in DSNMF, which can preserve the local geometric information of both the data space and the feature space, and fully excavate the potential data information. To solve the problem in the recently proposed DFSC [1], DSNMF adopts non-negative matrix factorization rather than self-representation matrices. Therefore, it can make the two non-negative matrix factors of the data space and the feature space update iteratively and interactively, which can give full play to the dual-graph model.

Considering that the previous clustering algorithms [22–28] are lack of discrimination, we combine K-means with LDA [29–31] and propose a feature selection based dual-graph sparse non-negative matrix factorization for local discriminative clustering (DSNMF-LDC). DSNMF-LDC not only has the advantages of the dual-graph model in the co-clustering algorithms, but also utilizes the feature selection which can remove some redundant and irrelevant features in the original features and select the discriminative features. Therefore, DSNMF-LDC can not only be robust to the noises, but also reduce the dimension of the data, save computing and storage resources. What's more, we utilize the local discriminative feature selection clustering after feature selection, which can greatly improve the clustering effectiveness and robustness.

Our main contributions are the following four aspects:

- We integrate NMF into feature selection rather than reduce the dimension by only NMF or feature selection, which can reduce the dimension of data as much as possible and efficiently deal with high-dimensional data.
- We combine the non-negative matrix factorization with the dual-graph (the data graph and the feature graph) model for feature selection, which can reduce the dimension of data as much as possible. The two non-negative matrix factors of the data space and the feature space can update iteratively and interactively, which can give full play to the dual-graph model.
- We exert $L_{2,1}$-*norm* constraint on the non-negative matrix factor of the feature space, which reflects the sparse self-representation information and the importance of selected features. What's more, it ensures the sparsity of the non-negative matrix factor of the feature space, which can simplify the calculation.
- We utilize the local discriminative feature selection clustering after feature selection, which can greatly improve the clustering effectiveness and robustness.

The rest of this paper is organized as follows: in Section 2, we introduce our feature selection based dual-graph sparse non-negative matrix factorization for local discriminative clustering (DSNMF-LDC) in detail. Extensive experiments and corresponding analyses are done in Section 3. In Section 4, we provide some conclusions and suggestions for the future work.

2. Feature selection based dual-graph sparse non-negative matrix factorization for local discriminative clustering (DSNMF-LDC)

2.1. Feature selection based dual-graph sparse non-negative matrix factorization (DSNMF)

2.1.1. Objective function

An advanced DFSC has been proposed in [1], which exerts the self-representation matrices of the data space and the feature space. However, it can only update by themselves and cannot affect each other, so that it cannot give full play to the dual-graph model. In order to solve this problem, we propose a feature selection based dual-graph sparse non-negative matrix factorization (DSNMF).

Non-negative matrix factorization can obtain the potential data information by decomposing the data matrix into non-negative matrix factors, which is a very effective method of the matrix approximation. We have a dataset $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \in \Re^{m\times n}$, where $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{im}]^T \in \Re^m$ is the ith vector, m is the number of the feature dimensions, n is the number of the simples. The purpose of non-negative matrix factorization is to decompose the data matrix into two non-negative matrix factors $\boldsymbol{P} = [\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_m]^T \in \Re^{m\times c}$ and $\boldsymbol{S} = [\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_n]^T \in \Re^{n\times c}$, where c is the clustering number, $\boldsymbol{P}$ is the non-negative matrix factor of the feature space