Pattern Recognition 55 (2016) 172–182

Contents lists available at ScienceDirect

Pattern Recognition

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

Global discriminative-based nonnegative spectral clustering

Ronghua Shang *, Zhu Zhang, Licheng Jiao, Wenbing Wang, Shuyuan Yang

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an 710071, China

ARTICLE INFO

Article history:
Received 20 December 2014
Received in revised form
17 October 2015
Accepted 29 January 2016
Available online 8 February 2016

Keywords:
Spectral clustering
Nonnegative matrix factorization (NMF)
Global discrimination information

ABSTRACT

Based on spectral graph theory, spectral clustering is an optimal graph partition problem. It has been proven that the spectral clustering is equivalent to nonnegative matrix factorization (NMF) under certain conditions. Based on the equivalence, some spectral clustering methods are proposed, but the global discriminative information of the dataset is neglected. In this paper, based on the equivalence between spectral clustering and NMF, we simultaneously maximize the between-class scatter matrix and minimize the within-class scatter matrix to enhance the discriminating power. We integrate the geometrical structure and discriminative structure in a joint framework. With a global discriminative regularization term added into the nonnegative matrix factorization framework, we propose two novel spectral clustering methods, named global discriminative-based nonnegative and spectral clustering (GDBNSC-Ncut and GDBNSC-Rcut) These new spectral clustering algorithms can preserve both the global geometrical structure and global discriminative structure. The intrinsic geometrical information of the dataset is detected, and clustering quality is improved with enhanced discriminating power. In addition, the proposed algorithms also have very good abilities of handling out-of-sample data. Experimental results on real word data demonstrate that the proposed algorithms outperform some state-of-the-art methods with good clustering qualities.

1. Introduction

Cluster analysis is an important part of data mining and pattern recognition [1,2], which is the problem of portioning the dataset into several categories according to certain similarity measure, so that data points belonging to the same class share high similarity, while the data points belonging to different classes have low similarity [3]. Clustering algorithms have been widely applied in many fields, such as image segmentation [4], genetic information analysis [5], document analysis [6], image retrieval [7], image compression [8], and so on.

Over the past decades, spectral clustering [9–14] has gained considerable attention from both the academic and the industrial communities. Compared with conventional clustering algorithms, spectral clustering has obvious advantages. It can converge to global optimum and that it performs well for the sample space of arbitrary shape, especially suitable for non-convex dataset [15]. Spectral clustering is based on algebraic graph theory, which treats data clustering problem as a graph partitioning problem [16]. It constructs an undirected weighted graph with each node corresponds to a data point, and the weight of the edge connecting the two nodes being the similarity value between the two points [17]. Then, using certain graph cut method, we divide the graph into connected components, which are called clusters. Typical graph cut methods include normalized cut (Ncut) [18], ratio cut (Rcut) [19], minimum cut (Mcut) [20] and min–max cut (MMcut) [21]. The optimal solution of graph partition can be obtained by minimizing or maximizing the objective function of the graph cut methods [22]. However, seeking the optimal solution of graph partition criteria is often NP-hard. Spectral clustering seeks to get the relaxation solution of graph cut objective function, which is an approximate optimal solution for graph partition. The basic idea is considering a continuous relaxation form of the original problem, turning to solve the eigenvalues and eigenvectors of the graph Laplacian matrix. In this paper, we only focus on spectral clustering approaches using Ncut and Rcut as objective functions.

Nonnegative matrix factorization (NMF) [23,24] is a typical method for dimensionality reduction and matrix factorization. NMF obtains a low-dimensional approximation of the original data matrix and gets a part-based representation of the data. The biggest difference between NMF and other matrix decomposition methods (such as SVD) is that the nonnegative constraints lead to the iterative multiplicative updating rules. By biological knowledge, we know that our brain has a part-based approach for recognition and understanding. The idea of NMF is consistent with our cognitive rules of the objective world [25,26]. Therefore, NMF has a clear physical meaning and strong interpretability.

* Corresponding author. Tel.: +86 29 88202279.
E-mail address: rhshang@mail.xidian.edu.cn (R. Shang).

http://dx.doi.org/10.1016/j.patcog.2016.01.035
0031-3203/

NMF is closely related to some algorithms in machine learning and pattern recognition communities. Probabilistic latent semantic indexing (PLSI) and NMF have been proven to be equivalent [27], although they are different methods, they optimize the same objective function. Ding et al. proved that kernel k-means can be treated as an NMF problem for symmetric matrix decomposition, and that NMF equals to Ncut spectral clustering [28]. It has also been pointed out that Laplace embedding is equivalent to Rcut spectral clustering [29]. In [29], the nonnegativity constraint is rigorously enforced, a nonnegative Laplacian embedding (NLE) approach is proposed and its links with NMF algorithm are demonstrated. In [28] and [29], symmetric NMF are involved. Different from [28] and [29], the data matrix itself is considered in [30]. Under proper conditions, Lu et al. demonstrate that a relaxed Rcut spectral clustering algorithm is equivalent to nonnegative factorization of the data matrix into the product a nonnegative matrix and another nonnegative matrix with orthogonal columns. Similarly, Ncut spectral clustering is also proven to be equivalent to nonnegative factorization of the normalized data matrix [30].

Under this equivalence, four algorithms: NSC-Ncut, NSC-Rcut, NSSC-Ncut and NSSC-Rcut are proposed in [30]. These four algorithms all consider the global manifold structure of a dataset, but they fail to consider the discriminative structure which reveals the intrinsic structure of the data distribution. We know that both manifold information and discriminant information are of great importance for clustering. We expect to preserve the discriminant information of a dataset in the learning process.

In order to capture the global discriminative information of the dataset, an intuitive approach is taking the class labels as prior knowledge in the learning process. However, in unsupervised clustering, it is infeasible to get the class labels in advance. Fortunately, in recent years, we have witnessed some progresses in employing discriminative structural information under the unsupervised learning paradigm [31–38].

Discriminative cluster analysis (DCA) [31] uses discriminative features for clustering rather than generative ones. Thus, clustering in the low dimensional discriminative space is more effective and computationally efficient than clustering in principal components space. In [32], the proposed discriminative k-means algorithm performs linear discriminant analysis (LDA) subspace selection and clustering simultaneously. In [33], both the local manifold structure and the global discriminant information are preserved simultaneously through manifold discriminant learning. In [34], the proposed local discriminative and global integration clustering algorithm (LDMGI) combines the local discriminative models and manifold structure for clustering. In [35], the discriminative information and geometrical information are characterized in a weighted feature space, which can well estimate the clustering structure of the data. In [36], a new Laplacian matrix was integrated into a spectral embedded clustering framework to capture local and global discriminative information for clustering. In [37], the global discriminative regularization term in is introduced, which provides more discriminative information to enhance clustering performance. In [38], an effective feature extraction method used discriminant analysis, which facilites the learning power of the method.

These algorithms use the global discriminative information, and make their performance to be improved. However, the general global discriminative model is used in linear cases, so these algorithms cannot effectively deal with the nonlinear data. Fortunately, this problem can be solved with the development of kernel tricks [39–44]. Kernel trick has been applied to many learning algorithms, such as the kernel principal component analysis (KPCA) [39], the kernel trick for support vector machines (SVMs) [40] and the kernelized LDA [41–44]. In [41], a nonlinear method based on Fisher's discriminant was proposed, which called kernel fisher discriminant (KFD). Fisher discriminant can be computed efficiently in feature space by using the kernel trick. So KFD can be used to handle nonlinear data, and also maintains the advantages of Fisher's discriminant analysis. The results show that KFD is competitive to other state-of-the-art methods. In [42], a method to deal with nonlinear discriminant analysis using kernel function operator was proposed. It is effective for both simulated data and alternate kernels. In [43], Liang et al. proposed a method to solve kernel Fisher discriminant analysis. This method is effective and feasible in dealing with handwritten numeral characters. In [44], the method of KFD was analyzed and a more transparent KFD algorithm was proposed, in which KPCA was first performed and then LDA was used for a second feature extraction. Simulation results on CENPARMI handwritten numeral database showed the effectiveness of this algorithm. Therefore, the kernelized global discriminative model can be used for nonlinear data effectively.

Inspired by these ideas, we integrate the global geometrical structure and the global discrimination structure in a joint unsupervised framework. We propose two novel spectral clustering algorithms named global discriminative-based nonnegative spectral clustering (GDBNSC-Ncut and GDBNSC-Rcut). The proposed approaches are expected to keep the connection between spectral clustering and NMF, and learn a compact data representation. This compact data representation can preserve not only the global geometric information but also has the global discriminant ability, both of which are crucial for effective clustering. Different from previous work [18,19,29,30], the proposed algorithms preserve both discriminative information and the geometrical information of the dataset, while still keeping the connection between NMF and spectral clustering.

We know that some former algorithms [28–30] just perform nonnegative matrix factorization of matrices to keep connection between spectral clustering and NMF. We go a step further by integrating discriminative information in the objective function to detect the intrinsic structure of the dataset.

It is worthwhile to highlight the main contributions of the proposed algorithms here:

1. The proposed methods do not only connect spectral clustering algorithm with NMF, but also characterize both the underlying global geometrical information and the global discriminative information of the dataset, and the proposed algorithms have good ability to handle out-of-sample data.
2. For the proposed algorithms, we give the objective functions, develop iterative multiplicative updating schemes, and analyze the convergence.
3. The remainder of this paper is organized as follows. In Section 2, we introduce some related work. In Section 3, we present the proposed algorithms, deduce iterative multiplicative updating rules, and then provide the convergence proof of the optimization scheme. Experimental part is presented in Section 4. Finally, some concluding remarks and several issues of future's work are given in Section 5.

2. Related works

In this section, we briefly review some recent work closely related to our algorithms.

2.1. Rcut spectral clustering

Let $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N] \in \mathbb{R}^{M \times N}$ denote the data matrix, $\boldsymbol{x}_i \in \mathbb{R}^M$ denotes the i-th data point, M is the dimensionality of original data, and N is the number of samples. The dataset is expected to group into K classes. We construct an undirected similarity graph $G = (V, E)$, where each node corresponds to a data