Pattern Recognition 45 (2012) 1205–1210

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

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

Graph optimization for dimensionality reduction with sparsity constraints

Limei Zhang [a,b], Songcan Chen [a,*], Lishan Qiao [b]

[a] Department of Computer Science and Engineering, Nanjing University of Aeronautics & Astronautics, 210016 Nanjing, PR China
[b] Department of Mathematics Science, Liaocheng University, 252000 Liaocheng, PR China

ARTICLE INFO

Article history:
Received 1 February 2011
Received in revised form
30 July 2011
Accepted 13 August 2011
Available online 23 August 2011

Keywords:
Dimensionality reduction
Graph construction
Sparse representation
Face recognition

ABSTRACT

Graph-based dimensionality reduction (DR) methods play an increasingly important role in many machine learning and pattern recognition applications. In this paper, we propose a novel graph-based learning scheme to conduct **G**raph **O**ptimization for **D**imensionality **R**eduction with **S**parsity **C**onstraints (GODRSC). Different from most of graph-based DR methods where graphs are generally constructed in advance, GODRSC aims to simultaneously seek a graph and a projection matrix preserving such a graph in one unified framework, resulting in an automatically updated graph. Moreover, by applying an l_1 regularizer, a sparse graph is achieved, which models the "locality" structure of data and contains natural discriminating information. Finally, extensive experiments on several publicly available UCI and face databases verify the feasibility and effectiveness of the proposed method.

1. Introduction

It is well known that dimensionality reduction (DR) has generally been used as a principled way to understand the high-dimensional data such as image, text and video sequence. Recently, graph-based DR methods become more and more popular in pattern recognition and machine learning fields, due to the fact that graph is a powerful tool to catch the structure information hidden in objects or data. In fact, recent research [1] claimed that most existing DR methods can fall into a graph embedding framework. The representatives have ISOMAP [2], LLE [3], Laplacian eigenmap [4] and locality preserving projections (LPP) [5], just to name a few. Under such a framework, one first constructs a graph from data in terms of some prior knowledge available, and then based on the constructed graph learns a projection matrix, which transforms the original high-dimension data into a lower dimensional space. Among them, graph construction is crucial since the performances of these algorithms depend heavily on how well the graph models the original data structure.

As a consequence, the methods for graph construction have been widely studied in recent years, although building a high-quality graph is still an open problem [6]. In general, most of the graph construction processes can be decomposed into two steps. Firstly, one constructs an adjacency graph by considering the samples as nodes and linking some of them with edges according to given rules such as *k*-nearest neighbors, ε-ball neighborhood and *b*-matching [4,5,7]. Secondly, a weight is assigned for each edge. The often-used weight assignment ways include Heat Kernel [4], Inverse Euclidean Distance [8] and Local Linear Reconstruction [3], etc. All these graph construction methods are quite flexible and can in principle be used for any graph-based learning algorithms including DR, spectral clustering and semi-supervised learning [1,7,9,10]. However, as pointed out in [10], there is potential need that graph should be appropriate for the subsequent learning task.

To establish an "appropriate" graph, Zhang et al. recently presented an algorithm called Graph-optimized Locality Preserving Projections (GoLPP) [11] for DR task, which optimizes graph and projections simultaneously in one single objective function. To the best of our knowledge, this is the first attempt to perform graph optimization during a specific DR process, rather than pre-define graph before DR as done in most of graph-based algorithms [1]. Despite GoLPP obtains empirical superiority to traditional LPP on some datasets, the graph resulted from GoLPP usually loses traditional sparsity even though a sparse initial graph is given in its iterative optimization.

To address this problem, in this paper, we propose a novel strategy to conduct **G**raph **O**ptimization for **D**imensionality **R**eduction with **S**parsity **C**onstraints (GODRSC). The proposed method not only shares the advantages of GoLPP with automatically adjustable graph, but also has some additional desirable characteristics:

1) The sparsity of graph is held by replacing the entropy regularizer in GoLPP with an l_1 norm minimization. As pointed out in [7], sparsification is important to graph since it can bring higher efficiency, better accuracy and robustness to noise.
2) Interestingly, with adjustable graph, GODRSC essentially provides an extension to the sparsity preserving projections (SPP) [12], a recently developed DR algorithm based on sparse

* Corresponding author. Tel.: +86 25 84896481x12221; fax: +86 25 84892400.
E-mail address: s.chen@nuaa.edu.cn (S. Chen).

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.08.015

representation (see Section 2.2 for more details). This establishes a natural link between GODRSC and SPP, which helps to give an intuitive explanation why and how the former might work well [12–14].

3) By solving the trace ratio problem directly, GODRSC avoids the nonuniqueness of the solutions involved in GoLPP. See Section 3 for more details.

The rest of the paper is organized as follows: Section 2 reviews three related DR algorithms, GoLPP, SPP and its orthogonalized extension. Section 3 introduces the GODRSC model and algorithm. In Section 4, some experimental results are presented. Finally, conclusions are drawn in Section 5.

2. Related works

2.1. Graph-optimized locality preserving projections (GoLPP)

Given a set of sample points $X=[x_1,x_2,\ldots,x_n]$, where $x_i \in R^D$, $i=1.2,\ldots,n$, we firstly review the objective function of LPP [5], which GoLPP is based on

$$\min_W \frac{\sum_{i,j=1}^n \|W^T x_i - W^T x_j\|^2 P_{ij}}{\sum_{i=1}^n D_{ii}\|W^T x_i\|^2}$$

where $W \in R^{D\times d}(d<D)$ is the projection matrix, $D_{ii}=\sum_{j=1}^n P_{ij}$, and $P=(P_{ij})_{n\times n}$ is the edge weight matrix of a neighbor graph, which has been specified before learning W. In contrast to LPP with such a pre-defined graph, GoLPP simultaneously completes graph optimization and projection learning within a unified objective function [11] below:

$$\min_{W,S_{ij}} \frac{\sum_{i,j=1}^n \|W^T x_i - W^T x_j\|^2 S_{ij}}{\sum_{i=1}^n \|W^T x_i\|^2} + \eta \sum_{i,j=1}^n S_{ij}\ln S_{ij}$$
$$\text{s.t.} \sum_{j=1}^n S_{ij}=1, i=1,\ldots,n$$
$$S_{ij}\geq 0, i,j=1,\ldots,n$$

which can in turn be rewritten as the following *trace ratio* form

$$\min_{W,S_{ij}} \frac{tr(W^T XLX^T W)}{tr(W^T XX^T W)} + \eta \sum_{i,j=1}^n S_{ij}\ln S_{ij}$$
$$\text{s.t.} \sum_{j=1}^n S_{ij}=1, i=1,\ldots,n$$
$$S_{ij}\geq 0, i,j=1,\ldots,n \tag{1}$$

where $S=(S_{ij})_{n\times n}$ is an unknown affinity weight matrix of graph, L is the graph Laplacian; $\sum_{i,j=1}^n S_{ij}\ln S_{ij}$ is an entropy regularization term with sum-to-one constraint $\sum_{j=1}^n S_{ij}=1$ and non-negative constraint $S_{ij}\geq 0$ for avoiding degenerate solution as well as endowing S_{ij} with probability meaning; η is a tradeoff parameter. According to Zhang et al. [11], the GoLPP model can be solved by alternating iteration and the iteration process is theoretically proved convergence. Finally its performance empirically outperforms LPP for visualization and classification tasks on a number of often-used public datasets, benefiting from the automatically optimized graph.

2.2. Sparsity preserving projections (SPP) and its orthogonalization

SPP [12] is an unsupervised DR algorithm based on graph construction by sparse representation. In particular, SPP firstly constructs a graph by representing each sample point x_i using as few sample points in $X\backslash\{x_i\}$ as possible. With different assumptions to noise, it can be cast into different l_1-minimization problems such as the following one:

$$\min_{S_i} \|S_i\|_1$$
$$\text{s.t.} \|x_i - XS_i\|^2 < \varepsilon$$
$$\sum_{j=1}^n S_{ij}=1 \tag{2}$$

where S_i is a column vector consisting of the representative coefficient of sample x_i,[1] and minimizing the l_1 norm aims to obtain a sparse solution; $\|x_i - XS_i\|^2$ is the error for reconstructing x_i. Naturally, the jth element S_{ij} in coefficient vector S_i can be used as the affinity weights between samples x_i and x_j, and thus SPP builds a graph $G=(X,(S_{ij})_{n\times n})$, which describes the sparse reconstructive relationship among the original samples.

Then, SPP seeks a projection matrix W best preserving the sparse graph above. Similar to NPE [15], a linear version of LLE [3], SPP does this by the following objective function:

$$\min_W \frac{\sum_{i=1}^n \|W^T(x_i - XS_i)\|^2}{\sum_{i=1}^n \|W^T x_i\|^2} \tag{3}$$

which is equivalent to the trace ratio problem:

$$\max_W \frac{tr(W^T XS_\beta X^T W)}{tr(W^T XX^T W)}$$

where$S_\beta = S+S^T - SS^T$. Similar to most trace ratio models [1,5,15], it can be approximately solved by generalized eigenvalue decomposition.

Furthermore, one can generalize SPP by introducing some priors or constraints to its model as in many linear DR algorithms such as LPP. In order to better discuss and validate the proposed GODRSC method later, here, we give an orthogonalized extension of SPP, and call it OSPP simply, which can be modeled just by imposing orthogonal constraint on projection matrix W. With the same notations as in Eq. (3), the model of OSPP is established by minimizing the objective defined as follows:

$$\min_W \frac{\sum_{i=1}^n \|W^T(x_i - XS_i)\|^2}{\sum_{i=1}^n \|W^T x_i\|^2}$$
$$\text{s.t. } W^T W = I$$

Similarly, we have its corresponding trace ratio form

$$\max_W \frac{tr(W^T XS_\beta X^T W)}{tr(W^T XX^T W)}$$
$$\text{s.t. } W^T W = I$$

With the orthogonal constraint, the OSPP model above can be solved *exactly* by many recently proposed algorithms [16–18], rather than *approximated* by the generalized eigenvalue problem as original SPP.

3. Graph optimization for dimensionality reduction with sparsity constraints

3.1. Motivations

Note that, in the GoLPP model (1), the maximum entropy term makes the edge weights of graph as uniform as possible, consequently incurring the loss of sparsity, which is a basic common merit in typical graph construction using k-NN and ε-ball, or l_1 regularization, etc. In fact, the resulted graph updating formulation (see Eq. (11) in Appendix) of GoLPP has shown that there exist nonzero edge weights in all the pairs of samples. On the

[1] It is worthwhile to point out that the ith entry in S_i is zero due to removing x_i from sample matrix X.