ARTICLE IN PRESS

Pattern Recognition ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Pattern Recognition

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

Image clustering based on sparse patch alignment framework

Jun Yu [a], Richang Hong [b,*], Meng Wang [b], Jane You [c]

[a] School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China
[b] Department of Computer and Information, Hefei University of Technology, Hefei, Anhui 230009, China
[c] Department of Computing, The Hong Kong polytechnic University, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 6 January 2014
Received in revised form
11 April 2014
Accepted 5 May 2014

Keywords:
Image clustering
Manifold learning
Sparse representation

ABSTRACT

Image clustering methods are efficient tools for applications such as content-based image retrieval and image annotation. Recently, graph based manifold learning methods have shown promising performance in extracting features for image clustering. Typical manifold learning methods adopt appropriate neighborhood size to construct the neighborhood graph, which captures local geometry of data distribution. Because the density of data points' distribution may be different in different regions of the manifold, a fixed neighborhood size may be inappropriate in building the manifold. In this paper, we propose a novel algorithm, named sparse patch alignment framework, for the embedding of data lying in multiple manifolds. Specifically, we assume that for each data point there exists a small neighborhood in which only the points that come from the same manifold lie approximately in a low-dimensional affine subspace. Based on the patch alignment framework, we propose an optimization strategy for constructing local patches, which adopt sparse representation to select a few neighbors of each data point that span a low-dimensional affine subspace passing near that point. After that, the whole alignment strategy is utilized to build the manifold. Experiments are conducted on four real-world datasets, and the results demonstrate the effectiveness of the proposed method.

1. Introduction

Recently, dimensionality reduction based techniques [1,2,37] show to have promising performance in image clustering and classification [41]. These methods assume images as high dimensional points distributing on or nearly on a low dimensional manifold. In past decades, many approaches [3–5,38] were proposed to build manifold by preserving the locality of manifold structure and the discrimination among different classes. The graph based methods have been widely adopted in many applications including multimedia information retrieval [6,29,30,42,43], document clustering [7], data visualization [8], and image matching [9]. Theoretically, graph is an efficient tool used in manifold learning [12,39,40] and semi-supervised learning [10,11].

Typical graph-based manifold learning methods include Locally Linear Embedding (LLE) [12], ISOMAP [19], Laplacian Eigenmaps (LE) [13], Local Tangent Space Alignment (LTSA) [14], and Locality Preserving Projections (LPP) [15]. LLE adopts linear coefficients to reconstruct a given measurement by its neighbors, and then explores a low-dimensional embedding, in which these coefficients are appropriate for reconstructing the local structure of the measurement. ISOMAP preserves global geodesic distances of all pairs of measurements. LE preserves relationships of proximity through an undirected weighted graph, which indicates neighbor relations of pairwise measurements. LTSA uses the local tangent information to describe the local geometry and this local tangent information is then aligned to provide a global coordinate. Recently, Zhang et al. [16] proposed another popular framework named Patch Alignment Framework (PAF), which unified manifold learning algorithms through two stages: local patch optimization and whole alignment. This framework explores that: (i) manifold algorithms are intrinsically different in the local patch optimization stage; and (ii) all algorithms share an almost-identical whole alignment stage. Guan et al. [33] applied patch alignment framework into the non-negative matrix factorization (NMF) related dimension reduction algorithms. Xia et al. [34] integrated with patch alignment framework with the multimodality learning. In these two methods, the neighbor number k in local patch construction is a parameter, which should be selected from a large range. Furthermore, graph-based semi-supervised learning methods, such as the popular Gaussian Laplacian based classification (Lap) [17] and the Transductive Classification via Local Learning algorithm (LL) [18], have attracted great interest in the research community. In graph-based semi-supervised learning methods, the construction of the Laplacian matrix, which

* Corresponding author: Tel.: +86 13866188972.
E-mail address: hongrc.hfut@gmail.com (R. Hong).

http://dx.doi.org/10.1016/j.patcog.2014.05.002
0031-3203/

Please cite this article as: J. Yu, et al., Image clustering based on sparse patch alignment framework, Pattern Recognition (2014), http://dx.doi.org/10.1016/j.patcog.2014.05.002

estimates the geometric structure of manifold, is critical to boost the performance.

Local methods, such as LLE, LE, LTSA and PAF, try to preserve local relationships among points by learning a set of weights between each point and its neighbors. Global methods, such as ISOMAP, try to preserve local and global relationships among all data points. All these methods estimate the manifold of the data using several eigenvectors of a matrix related to the learned weights between pairs of points. Therefore, a proper selection of the neighborhood size adopted to construct the neighborhood graph is important for both local and global methods. In detail, a small neighborhood size may not capture sufficient information about manifold geometry estimation, while a large neighborhood size could break the rules used to capture information about the manifold. In addition, the density of data points' distribution may be different in different regions of the manifold. Therefore, a fixed neighborhood size may be inappropriate in building the manifold.

In this paper, we propose a novel algorithm, called sparse patch alignment framework (SPAF), for the embedding of data lying in multiple manifolds. Specifically, we assume that for each data point there exists a small neighborhood in which only the points that come from the same manifold lie approximately in a low-dimensional affine subspace. Based on the patch alignment framework, we propose an optimization strategy for constructing local patches, which adopt sparse representation to select a few neighbors of each data point that span a low-dimensional affine subspace passing near that point. Hence, a few nonzero elements of the solution indicate the points that are on the same manifold, hence they can be used for building local patches. After that, the whole alignment strategy is utilized to build the manifold. Embedding of the data into lower dimensions follows by taking the eigenvectors of the matrix. In summary, contributions of this paper are as follows:

- First, a novel framework of manifold construction is proposed. Sparse representation is used in local patch construction, and whole alignment is adopted to obtain the global manifold. This framework can be used for different applications including image clustering and classification.
- Second, sparse representations of each data point can reflect the intrinsic dimensionality of the underlying manifold. Besides, the proposed SPAF has only one free parameter that results in a stable embedding, and allows robust automatic selection of neighbors.

The rest of this paper is organized as follows. Section 2 introduces the related work. The algorithm of sparse patch alignment framework is proposed in Section 3. Section 4 provides experiments of image clustering on four image datasets. The application of the framework for image classification is introduced in Section 5. Conclusions are drawn in Section 6.

2. Related works

In many real-world applications, the samples distribute in multiple manifolds with different dimensions [20]. In order to build a low-dimensional embedding of the data, the samples should be first separated into different clusters, and then a low-dimensional representation for the samples in each cluster should be constructed. Since the manifolds can be very close to each other and they can have arbitrary dimensions, curvature and sampling, the manifold clustering and embedding problem [20] is very challenging. Wu et al. [21] adopted an agglomerative process to separate independent manifolds. Initially, data samples are segmented into different groups according to a threshold. Subspace is then fit to each group, and two groups are merged when the distance between their subspaces is below a threshold. Generalized PCA (GPCA) [22] is an algebraic geometric method for clustering data lying in linear subspaces. The main idea behind GPCA is that one can fit a union of subspaces with a set of polynomials, whose derivatives at a point give a vector normal to the subspace containing that point. Specifically, the first step of GPCA projects the samples onto a subspace. The second step is to fit a homogeneous polynomial of the projected samples. The last step is to compute the normal vectors from the vector of coefficients. The segmentation of the samples is then obtained by grouping these normal vectors using possible techniques.

The above mentioned methods take advantage of the global linear relations among samples in the same subspace. Therefore, they cannot handle nonlinear manifolds. Besides, these methods adopt complicated procedures to achieve manifold clustering, which cannot be conveniently used in building novel graph-based learning algorithms.

3. Sparse patch alignment framework

Patch Alignment Framework (PAF) [16] integrates popular manifold learning algorithms, e.g., LLE, ISOMAP, LTSA and LPP. It obtains two stages: local patch optimization and whole alignment, which can be utilized to design manifold learning algorithms with specific objectives. Table 1 lists important notations used in this paper.

Local patch optimization: For a given sample $\vec{\mathbf{v}}_i$ in a dataset, PAF constructs a local patch $\mathbf{V}_i = [\vec{\mathbf{v}}_i, \vec{\mathbf{v}}_{i^1}, \ldots, \vec{\mathbf{v}}_{i^k}]$, wherein $\vec{\mathbf{v}}_{i^1}$, …, $\vec{\mathbf{v}}_{i^k}$ are k nearest samples of $\vec{\mathbf{v}}_i$. For each patch $\mathbf{V}_i$, the corresponding low-dimensional representation is denoted by $\mathbf{H}_i = [\vec{\mathbf{h}}_i, \vec{\mathbf{h}}_{i^1}, \ldots, \vec{\mathbf{h}}_{i^k}]$. Then, the optimization of local patches [16] encodes the proximity information over the local patch $\mathbf{H}_i$ by preserving the distances between $\vec{\mathbf{h}}_i$ and its k nearest samples as small as possible. The optimization is conducted as $\min_{H_i} tr(\mathbf{H}_i \mathbf{L}_i \mathbf{H}_i^T)$, where $tr(\bullet)$ is the trace operator, and $\mathbf{L}_i$ varies with different manifold learning algorithms to capture the local geometry of the patch.

Table 1
Important notations used in this paper.

Notations	Descriptions
$\mathbf{V}_i = [\vec{\mathbf{v}}_i, \vec{\mathbf{v}}_{i^1}, \ldots, \vec{\mathbf{v}}_{i^k}]$	A local patch with k nearest samples of $\vec{\mathbf{v}}_i$
$\mathbf{H}_i = [\vec{\mathbf{h}}_i, \vec{\mathbf{h}}_{i^1}, \ldots, \vec{\mathbf{h}}_{i^k}]$	The low-dimensional representation
$\mathbf{L}_i$	Captures the local geometry of the local patch
$\mathbf{S}_i$	The selection matrix
$\{\Psi_l\}_{l=1}^n$	n different manifolds with varied dimensions $\{d_l\}_{l=1}^n$
$\mathbb{N}_i$	Represent the set of all samples in B_i without $\vec{\mathbf{v}}_i$
$\mathbf{P}_i$	It is a positive-definite diagonal matrix recording the adjacent relationships between $\vec{\mathbf{v}}_i$ and other points
$\mathbf{U}$	The projection matrix