Neurocomputing 72 (2009) 1719–1731

Contents lists available at ScienceDirect

Neurocomputing

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

Nonlinear dimensionality reduction with relative distance comparison

Chunxia Zhang [a], Shiming Xiang [b,*], Feiping Nie [c], Yangqiu Song [c]

[a] Software School, School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China
[b] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[c] Department of Automation, Tsinghua National Laboratory for Information Science and Technology (TNList), Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 26 October 2007
Received in revised form
27 July 2008
Accepted 12 August 2008
Communicated by T. Heskes
Available online 11 September 2008

Keywords:
Nonlinear dimensionality reduction (NLDR)
Relative distance comparison
Semi-definite programming

ABSTRACT

This paper proposes a new algorithm for nonlinear dimensionality reduction. Our basic idea is to explore and exploit the local geometry of the manifold with relative distance comparisons. All such comparisons derived from local neighborhoods are enumerated to constrain the manifold to be learned. The task is formulated as a problem of quadratically constrained quadratic programming (QCQP). However, such a QCQP problem is not convex. We relax it to be a problem of semi-definite programming (SDP), from which a globally optimal embedding is obtained. Experimental results illustrate the validity of our algorithm.

1. Introduction

In many areas of natural and social sciences, one is often confronted with intrinsically low-dimensional data points which lie in a very high-dimensional observation space. An example might be a set of images of an individual's face observed under different poses. If there are $n \times n$ grayscale pixels totally, then each face image yields a data point in n^2-dimensional space. But the intrinsic dimensionality of the space of all these images only equals to the number of the pose parameters. Reducing the dimensionality of such data is needed in many applications, ranging from image compression [34] to data visualization [27,4] and other tasks including recognition [33,35], vision computing [16,11], and so on [26,5]. Generally, the motivation behind dimensionality reduction is to discover a lower-dimensional structure in high-dimensional data without significant loss of information. Linear methods, such as principal component analysis (PCA) [12] and classical multidimensional scaling [6] are popularly used to perform the task of dimensionality reduction. Such linear methods are able to discover the lower-dimensional structure of data lying on or nearly lying on a linear space. However, in many cases lower-dimensional structure hidden in the data is nonlinear and directly using those linear methods may generate unsatisfactory results.

Recently, many nonlinear dimensionality reduction (NLDR) algorithms [15] have been developed under the assumption that the data points are sampled from an underlying manifold embedded in a high-dimensional Euclidean space. The two well-known algorithms are Isomap [25] and locally linear embedding (LLE) [18]. Based on the multidimensional scaling algorithm, Isomap attempts to preserve globally the geodesic distances between any pair of data points. LLE tries to discover the nonlinear structure by exploiting the local geometry of the data. Later, different manifold learning algorithms have been proposed, such as manifold charting [3], Laplacian eigenmap (LE) [1], Hessian LLE (HLLE) [7], local tangent space alignment (LTSA) [36], maximum variance unfolding (MVU) [29,28], conformal eigenmap [21], and other extensive work [15,31,22,10,24,13,14], etc. Most NLDR algorithms can be considered into a common framework of *thinking globally and fitting locally*, in which the locally geometrical information is collected together to obtain a global optimum [19].

Locally geometrical information of data is explored and exploited in different ways. In LLE [18], each data point is linearly reconstructed with its neighbors and such a linear representation is maintained in a lower-dimensional space. LE calculates the similarity of any pair of neighboring data points to define the graph Laplacian [1]. HLLE, LTSA and LSE explore the local relations between neighboring data points in tangent spaces. Local coordinates are mapped, linearly [36,24] or nonlinearly [31], to the global coordinate system with lower dimensionality. In contrast, MVU [29,28] and conformal eigenmap [21] utilize the locally geometrical relations in a straightforward way. In MVU, Euclidean distances between neighboring data points are globally

* Corresponding author. Tel.: +86 10 627 96 872; fax: +86 10 627 86 911.
E-mail address: smxiang@gmail.com (S. Xiang).

0925-2312/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2008.08.003

preserved in lower-dimensional space. Such an idea is extended in conformal eigenmap by preserving angle information and similar results can be obtained.

This paper uses relative distance comparisons to explore the local geometrical relations between data points. As a kind of side information, employing *relative comparison* is not a new idea in machine learning. Actually it is used to learn distance metrics for data clustering and classification with "*A is more similar to B than A is to C*" [20,17]. In addition, as supervised information, such a formulation is translated as "*the distance from A to B is not greater than that from A to C*" in manifold alignment [32]. In this paper, all such relative comparisons derived in each neighborhood on the manifold are enumerated and maintained in lower-dimensional manifold to be learned. The task is formulated as a problem of quadratically constrained quadratic programming (QCQP) [2]. However, such an optimization problem is nonconvex. To remedy this drawback, we perform a semi-definite slack and convert the source QCQP problem to be a problem of semi-definite programming (SDP) [2], in which a global embedding is finally obtained.

One advantage of relative distance comparison is that it can be easily specified not only in the local neighborhoods but also in the global region. This yields a mechanism to integrate prior knowledge or supervised information into manifold learning [32]. Performances show that adding a few relative comparisons about the global manifold structure may significantly change the learned shape. In contrast, most manifold learning algorithms could not be directly extended to integrate the global information in such a straightforward way without changing the nature of the optimization model.

In distance preserving framework [29,28],global structure can also be specified with distances between non-neighboring data points. However, supplying such distances is not an easy task. One reason is that the values in equality constraints should be carefully input to avoid conflicts between equalities. Another reason is that geodesic distance or manifold distance should be considered for the non-neighboring data points since the commonly used Euclidean distance metric can only be suitable for neighboring data points. On the one hand, calculating geodesic distances usually needs much time. On the other hand, the obtained geodesic distance may not reflect the true distance, specially when the data are sparsely sampled from the manifold or the topological structure of the manifold is not convex. Such drawbacks can be easily avoided with relative comparisons.

The remainder of this paper is organized as follows: Section 2 will develop the optimization model for NLDR problem. Section 3 will discuss how to solve the model with SDP. The extension of the model for using global relative distance comparisons is also presented in this section. Section 4 gives the algorithm. We report the experimental results in Section 5 and draw conclusions in Section 6.

2. The model

The NLDR problem can be formulated as follows. Given a set of n scattered data points $\mathbf{x}_i \in \mathbb{R}^m$ lying on a manifold M embedded in a m-dimensional Euclidean space. The goal is to invert an underlying generative model $\mathbf{x} = f(\mathbf{y})$ to find the corresponding lower-dimensional parameters (embedding coordinates) $\mathbf{y}_i \in \mathbb{R}^d$ such that $\mathbf{x}_i = f(\mathbf{y}_i)$, that is, construct $\mathscr{Y} = \{\mathbf{y}_i\}_{i=1}^n$ from $\mathscr{X} = \{\mathbf{x}_i\}_{i=1}^n$.

Differing from those with locally linear reconstruction and tangent space representation [18,7,36,31], here we use purely geometrical representations to explore the relations between data points.

For each data points $\mathbf{x}_i \in \mathscr{X}$ $(i = 1, \ldots, n)$, denote its neighborhood by $\mathscr{N}_i$, which contains k nearest neighbors of $\mathbf{x}_i$ obtained with Euclidean distance metric. Further let $\mathscr{N}_i = \{\mathbf{x}_{i_j}\}_{j=1}^k$, in which subscript i_j stands for an index and $i_j \in \{1, 2, \ldots, n\}$. Now we can enumerate all the triples $\mathscr{S}_i = \{(i, i_j, i_k) \mid \mathbf{x}_{i_j}, \mathbf{x}_{i_k} \in \mathscr{N}_i\}$ and each triple (i, i_j, i_k) corresponds to a relative distance comparison:

$$\|\mathbf{x}_i - \mathbf{x}_{i_j}\| \leqslant \|\mathbf{x}_i - \mathbf{x}_{i_k}\| \tag{1}$$

Here $\|\cdot\|$ stands for the L2-norm. Eq. (1) indicates that the distance from $\mathbf{x}_i$ to $\mathbf{x}_{i_j}$ is not greater than that from $\mathbf{x}_i$ to $\mathbf{x}_{i_k}$. Such a distance comparison reflects a weak geometry between data points.

Now we hope such a relative comparison is also maintained in lower-dimensional intrinsic space. It follows

$$\|\mathbf{y}_i - \mathbf{y}_{i_j}\| \leqslant \|\mathbf{y}_i - \mathbf{y}_{i_k}\| \tag{2}$$

Obviously, only relative comparisons cannot yield a unique solution. Additional constraints should be introduced to develop the optimization model.

First, the lower-dimensional coordinates $\mathscr{Y} = \{\mathbf{y}_i\}_{i=1}^n$ should be embedded into a definitely specified place. Such a place can be selected near the coordinate origin:

$$\sum_{i=1}^{n} \mathbf{y}_i = \mathbf{0} \tag{3}$$

Second, to avoid generating zero solutions, a distance guard should be introduced. It can be selected as the minimum distance among all the pairs of neighboring data points. Without loss of generality, suppose $\mathbf{x}_1$ and $\mathbf{x}_2$ can give the minimum distance. Then we have

$$\|\mathbf{y}_1 - \mathbf{y}_2\| = c \tag{4}$$

Here c equals to the distance from $\mathbf{x}_1$ to $\mathbf{x}_2$. Note that there only exists a scaling difference if c is specified as any other positive number. Therefore, we simply take $c = 1$ in this paper.

Unfortunately, only satisfying the constraints in (2)–(4) can still not generate a unique solution. Fig. 2 shows an example. In Fig. 2(a) five source data points A, B, C, D and E are sampled from a circle. This is a one-dimensional manifold with angle parameters as the intrinsic parameters. Given the distance from A to B as a distance guard, the line segments "ABCDE", "$\mathrm{ABC_1D_1E_1}$" and "$\mathrm{ABC_2D_2E_2}$" in Fig. 1(b) are all feasible solutions. Among those feasible solutions, we select that with maximum variance since with this solution the manifold is largely unfolded. This solution is just the one as "ABCDE" in Fig. 1(b), which shows a perfect one-dimensional embedding.

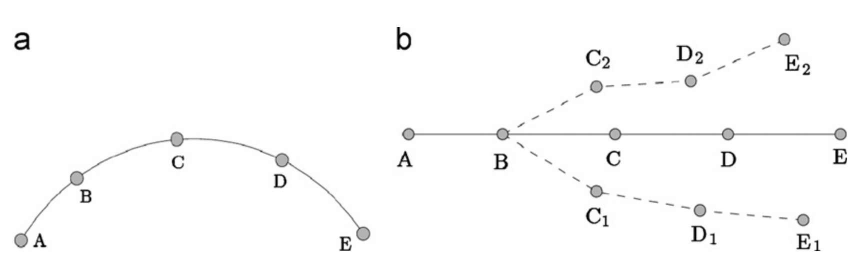

Fig. 1. (a) Five source data points sampled from a circle; (b) feasible solutions satisfying the constraints in Eqs. (2)–(4).