Neurocomputing 136 (2014) 7–13

Contents lists available at ScienceDirect

Neurocomputing

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

Letters

Real local-linearity preserving embedding

Jing Wang

The School of Computer Science and Technology, Huaqiao University, Xiamen 361021, PR China

ARTICLE INFO

Article history:
Received 9 July 2013
Received in revised form
27 October 2013
Accepted 15 January 2014
Communicated by Zhi Yong Liu
Available online 15 February 2014

Keywords:
Manifold learning
Local geometry
Locally linear embedding
Wrong neighbors

ABSTRACT

In recent years, manifold learning methods have aroused a great interest in the machine learning community. A key issue that determines the effectiveness of the manifold learning methods is how to accurately capture the local geometry of the low-dimensional manifold. However, most of the manifold learning algorithms cannot exploit the real local geometry if the neighbors for each sample point are not correctly selected. In this paper, we address this problem in the context of locally linear embedding (LLE). A new local optimization model is proposed to find the local weights that can represent the real local manifold geometry when the neighborhoods contain wrong neighbors. A new algorithm called real local-linearity preserving embedding (RLLPE) is then proposed by preserving the exploited real local geometry. We demonstrate the improvement and efficiency of RLLPE using both synthetic and real-word data.

1. Introduction

The problem of dimensionality reduction, i.e., the transformation of high-dimensional data into meaningful low-dimensional features, has aroused a great deal of interest in many research fields, such as pattern recognition [6], data mining [23], image processing [21], and computer vision [11]. Recently, there have been much effort in developing effective and efficient algorithms for learning meaningful nonlinear structure from the high-dimensional data. These algorithms which include isometric mapping (Isomap) [17], locally linear embedding (LLE) [15], Laplacian eigenmaps (LE) [1], local tangent space alignment (LTSA) [25], and so on, have some advantages over traditional linear dimensionality reduction techniques.

Most of the existing manifold learning algorithms fail in learning the low-dimensional structure if the neighbors of each point are not correctly selected. This is because that the local linear structure determined by the local optimization model cannot represent the real local geometry of the manifold when the neighborhoods contain wrong neighbors. For example, ISOMAP may estimate wrong geodesic distances because of wrong neighbors. In LTSA, wrong neighbors may lead to large bias in estimating the local tangent space. In LLE, the calculated reconstruction weights of the wrong neighbors may not represent the real local geometry. In this paper, we address this problem in the context of LLE.

LLE is one of the most wildly used manifold learning methods because of its simple geometric intuitions, straightforward implementation, and global optimization. However, it is also reported that LLE may not be stable since the constrained least squares (LS) problem involved for determining the local weights may be ill-conditioned [19]. In the literature, there are two strategies for improving the stability of the local geometry. A commonly used strategy is to introduce a regularization parameter for the constrained LS problem [5,14,16]. Another strategy is to explore the local geometry by multiple weight vectors or solving new constrained LS problem [7,19]. Despite the appealing properties of these extensions in improving the stability of LLE, they have a limited effectiveness when the neighborhoods contain wrong neighbors. More detailed introductions on the extensions of LLE are proposed in Section 4.

In this paper, we construct a new LS optimization model to find the local weights that exploit the real local geometry. We try to penalize the weights of wrong neighbors such that the undesirable effect of the wrong neighbors on the reconstruction can be reduced. A new algorithm called real local-linearity preserving embedding (RLLPE) is then proposed by preserving the exploited real local geometry.

The rest of this paper is organized as follows. In Section 2, we give a quick review of LLE, and discuss some of its failure modes. The real local-linearity preserving embedding algorithm (RLLPE) is then presented in Section 3. After that, the related work will be introduced in Section 4. We will give numerical experiments in Section 5 to show the effectiveness of RLLPE. Some conclusion remarks are given in Section 6.

2. A brief review of locally linear embedding

In this section, we first outline the basic steps of LLE and illustrate its failure mode using two simple examples. Given a data

E-mail address: wroaring@yahoo.com.cn

http://dx.doi.org/10.1016/j.neucom.2014.01.040
0925-2312

set $X=[x_1,\ldots,x_N]$ with $x_i \in \mathcal{R}^m$, sampled (possibly with noise) from a d-dimensional manifold ($d<m$), LLE proceeds in the following steps.

(1) *Setting local neighborhoods*: For each x_i, $i=1,\ldots,N$, determine the neighbor set $\mathcal{N}_i=\{x_{i_1},\ldots,x_{i_k}\}$ of its neighbors.

(2) *Extracting local geometry*: For each x_i, the local geometry of its neighborhood $\mathcal{N}_i$ is determined by solving the constrained least squares problem

$$\min_{w_{ji}, j\in J_i} \|x_i-\sum_{j\in J_i} w_{ji}x_j\|^2, \quad \text{s.t.} \quad \sum_{j\in J_i} w_{ji}=1, \tag{1}$$

where $J_i=\{i_1,\ldots,i_k\}$ is the index set of the neighbors.

(3) *Embedding global coordinates*: Map $\{x_1,\ldots,x_N\}$ to $\{t_1,\ldots,t_N\}$ in a d-dimensional space that preserves the local reconstruction properties totally

$$\min_{T=[t_1,\ldots,t_N]} \sum_i \|t_i-\sum_{j\in J_i} w_{ji}t_j\|^2, \quad \text{s.t.} \quad TT^T=I_d, \tag{2}$$

where I_d is a $d\times d$ identity matrix.

Denote $w_i=[\ldots,w_{ji},\ldots]^T$, $j\in J_i$ and $G_i=[\ldots,x_j-x_i,\ldots]$, $j\in J_i$. The constrained LS problem (1) can be rewritten as

$$\min_{w_i} \|G_iw_i\|^2, \quad \text{s.t.} \quad \mathbf{1}_k^Tw_i=1, \tag{3}$$

where $\mathbf{1}_k$ is the k-dimensional column vector with all ones. Employing Lagrange theorem for minimizing (3), the optimal weight vector can be given by

$$G_i^TG_iy_i=\mathbf{1}_k, \quad w_i=y_i/\mathbf{1}_k^Ty_i. \tag{4}$$

If $G_i^TG_i$ is singular or nearly singular, it is not stable to obtain the weights from (4). And it is suggested to solve the regularized linear system replaced

$$(G_i^TG_i+\gamma I_k)y_i=\mathbf{1}_k, \quad w_i=y_i/\mathbf{1}_k^Ty_i. \tag{5}$$

For example, it is proposed in [16] to set $\gamma=(\triangle^2/k)\|G_i\|_F^2$ with $\triangle=0.1$.

Generally, LLE works well if the neighbors for each sample point are correctly selected, i.e., the selected neighbors can reflect the local geometric structure of the manifold. However, if the neighborhoods contain wrong neighbors, LLE may fail even if the regularization parameter γ is well selected. In the rest of this section, we give two failure examples of LLE.

Example 1. The data set is generated as $x_i=[t_i, \ 10e^{-t_i^2}]^T$, $i=1\ldots,180$, where $t_i\in[-6, \ 6]$ are equally spaced. The sample points are of a highly nonuniform density since the curvature of the 1-D curve changes from 0 to 20 over $t\in[-6, \ 6]$.

Example 2. We generate the original data points as $x_i=[\sin(t_i), \ \cos(t_i), 3t_i]^T$, $i=1,\ldots,400$ with t_i equally spaced in $[0,4\pi]$. Then we generate $p=250$ noisy points as

$$y_j=x_{i_j}+0.5\text{randn}(3,1), \quad j=1,\ldots,p,$$

where x_{i_j} are randomly selected from the original data set and *randn* is Matlab's standard normal distribution.

For the above two data sets, the neighbor sets determined by k-NN or ϵ-neighborhood methods may contain wrong neighbors; see the left column in Fig. 1. LLE fails on these data sets. As it can be seen in Fig. 1, the computed coordinates by LLE with different γ cannot recover the arc-length coordinates. We remark that LLE combined with the regularization parameter selection methods proposed in [5,14,16] also fails on the above two data sets (the results are similar but are not reported here for brevity). This is

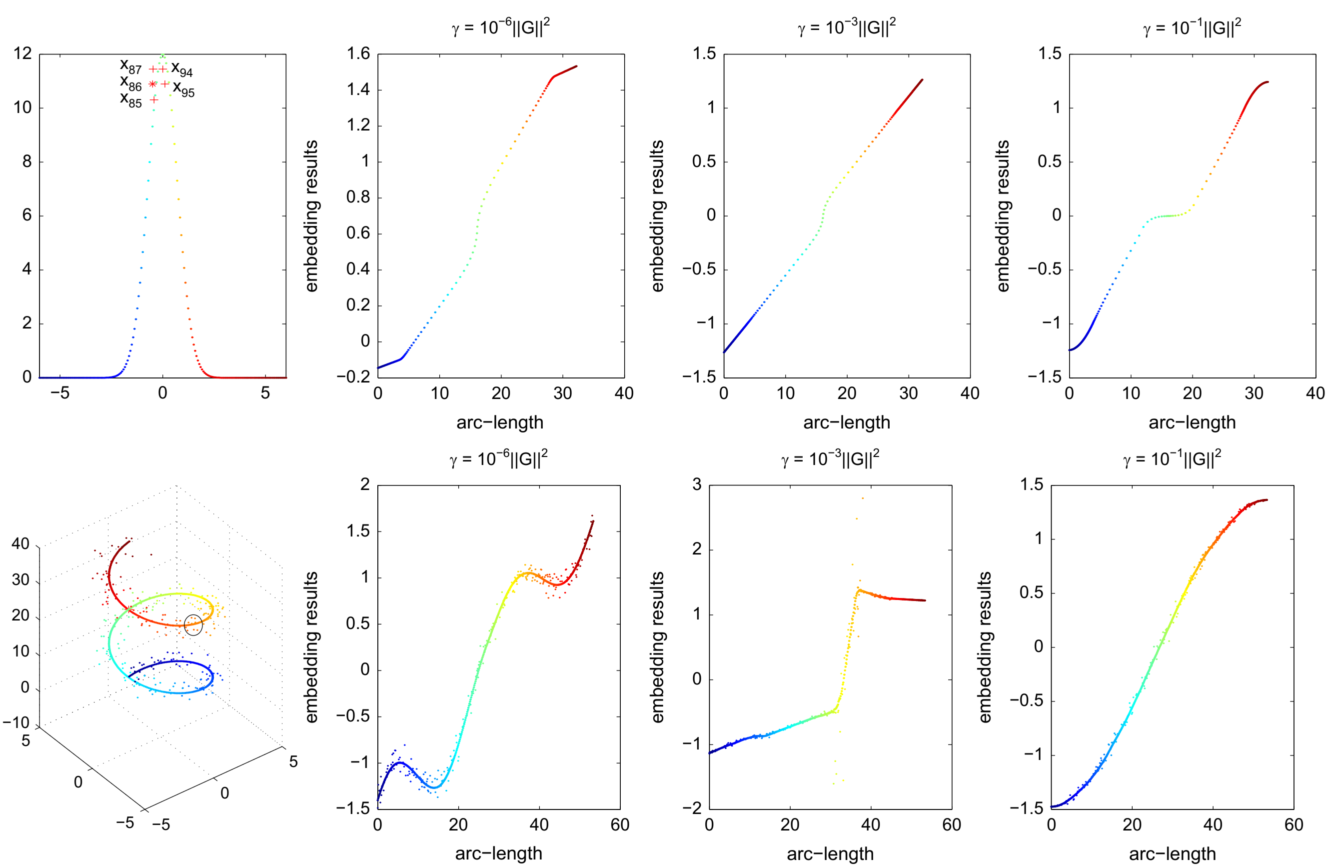

Fig. 1. Examples 1 and 2: from left to right – the data set, the coordinates of LLE with $\gamma=10^{-6}\|G\|_F^2, 10^{-3}\|G\|_F^2$ and $10^{-1}\|G\|_F^2$ on the curve with highly varying curvature (top row) and the curve with noise (bottom row) versus the arc length.