

Research Article

Dimensionality Reduction by Weighted Connections between Neighborhoods

Fuding Xie,^{1,2} Yutao Fan,¹ and Ming Zhou¹

¹ School of Urban and Environmental Science, Liaoning Normal University, Liaoning, Dalian 116029, China

² Academy of Mathematics and System Sciences, Chinese Academy of Science, Beijing 100080, China

Correspondence should be addressed to Fuding Xie; xiefd@lnnu.edu.cn

Received 6 January 2014; Accepted 24 February 2014; Published 22 April 2014

Academic Editor: Caihong Li

Copyright © 2014 Fuding Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Dimensionality reduction is the transformation of high-dimensional data into a meaningful representation of reduced dimensionality. This paper introduces a dimensionality reduction technique by weighted connections between neighborhoods to improve K -Isomap method, attempting to preserve perfectly the relationships between neighborhoods in the process of dimensionality reduction. The validity of the proposal is tested by three typical examples which are widely employed in the algorithms based on manifold. The experimental results show that the local topology nature of dataset is preserved well while transforming dataset in high-dimensional space into a new dataset in low-dimensionality by the proposed method.

1. Introduction

Real-world data, such as speech signals, digital photographs, or fMRI scans, usually has a high dimensionality. In order to handle such real-world data adequately, its dimensionality needs to be reduced. Ideally, the reduced representation should have a dimensionality that corresponds to the intrinsic dimensionality of the data. Fortunately, in many of those applications, all of the components of those high-dimensional data vectors are not independent of each other and in many cases the data points can be considered as lying on or close to a low-dimensional nonlinear manifold embedded in a high-dimensional space. Dimensionality reduction methods can also be thought of as a principled way to understand the high-dimensional data.

Mathematically, the problem of dimensionality reduction can be described as follows briefly. Assume we have a dataset with n samples in D -dimensional space. Assume further that this dataset has intrinsic dimensionality d (where $d < D$ and often $d \ll D$). The intrinsic dimensionality means that the points in dataset are lying on or near a manifold with dimensionality d that is embedded in the D -dimensional space. The manifold may be non-Riemannian because of discontinuities. Dimensionality reduction techniques transform dataset with

dimensionality D into a new dataset with dimensionality d , while retaining the geometry of the data as much as possible [1].

A large number of nonlinear techniques for dimensionality reduction have been proposed in the past decades. Among them, principal component analysis (PCA) [2], linear discriminant analysis (LDA) [3], and multidimensional scaling (MDS) [4] are three most classical techniques for dimensionality reduction. PCA finds a low-dimensional embedding of the data points that best preserves their variance as measured in the high-dimensional input space. Classical MDS finds an embedding that preserves the interpoint distances, equivalent to PCA when those distances are Euclidean. LDA searches the projection axes on which the not-same-class points are far from each other while requiring the same-class points to be close to each other. In 2000, two novel methods for manifold learning, isometric feature map method (Isomap) [5] and the locally linear embedding method (LLE) [6], have drawn great interests. Unlike other nonlinear dimension reduction methods, both LLE and Isomap methods emphasize simple algorithmic implementation and avoid nonlinear optimization formulations that are prone to local minima [7]. In Isomap, the geodesic distances between the data points are computed by constructing a neighborhood graph G ,

in which every data point is connected with its k nearest neighbors. The geodesic distances between all data points in dataset are computed, thereby forming a pairwise geodesic distance matrix. The low-dimensional representations are obtained by applying classical MDS on the resulting pairwise geodesic distance matrix. In LLE, the local properties of the data manifold are constructed by writing the high-dimensional data points as a linear combination of their nearest neighbors. In the low-dimensional representation of the data, LLE attempts to retain the reconstruction weights in the linear combinations as good as possible. van der Maaten et al. [1] presented a review and systematic comparison of a variety of nonlinear dimensionality reduction techniques which include Kernel PCA, maximum variance unfolding, diffusion maps, Laplacian eigenmaps, local tangent space analysis, Sammon mapping, multilayer autoencoders, locally linear coordination, manifold charting, Isomap, and LLE. The performances of these nonlinear techniques are investigated on artificial and natural tasks.

Recently, a great many methods of dimensionality reduction methods based on manifold learning have been introduced and some classical techniques have been further improved [8–15]. In [8], Hu et al. proposed a new dimensionality reduction algorithm called discriminant multidimensional mapping (DMM), which combines the advantages of multidimensional scaling (MDS) and LDA. DMM is effective for small sample datasets with high dimensionality. A dimensionality reduction technique, named orthogonal isometric projection (OIP), is proposed in [9]. In contrast with Isomap, which learns the low-dimension embedding and solves problem under the classic multidimensional scaling (MDS) framework, they consider an explicit linear projection by capturing the geodesic distance, which is able to handle new data straightforwardly and leads to a standard eigenvalue problem. To address the problem of finding a template function that represents the common pattern of a sample of curves, a novel algorithm based on a robust version of the isometric featurizing mapping (Isomap) algorithm is developed by Dimeglio et al. [11]. H. Choi and S. Choi [12] presented a robust kernel Isomap which relates the Isomap to Mercer kernel machines, so that the generalization property naturally emerges, through kernel principal component analysis. However, it is still a challenging problem to further propose a novel method to reduce the dataset in high-dimensional space into a desired dataset in low-dimensionality; meanwhile, the local topology nature of dataset is preserved perfectly.

Motivated by the Isomap and LLE dimensionality reduction methods proposed by Tenenbaum et al. [5] and Roweis and Saul [6], respectively, we in this paper improve the K -Isomap algorithm by local weighted connection between neighbors. The nearness between neighbors to a data point can be reflected well by the proposed technique. The desired dimensionality reduction results have been obtained on the three classical examples by the proposal.

The paper is organized as follows. Section 2 introduces classical MDS and Isomap dimensionality reduction techniques. An improvement of K -Isomap algorithm is depicted in Section 3. Section 4 shows the validity of the proposal by

two well-known datasets which are widely used by a large number of dimensionality reduction techniques. The conclusion is given in the last section.

2. MDS and Isomap Techniques

2.1. Classical MDS. MDS [4, 8, 16] is one of the global nonlinear techniques for dimensionality reduction which attempts to preserve global properties of the data. It maps the high-dimensional data representation to a low-dimensional representation while retaining the pairwise distances between the data points as faithfully as possible.

Suppose for a set of N points $X = \{x_1, x_2, \dots, x_N\}$, $x_i \in R^m$. The pairwise Euclidean distance between the data points x_i and x_j is

$$d_{ij} = d(x_i, x_j) = \|x_i - x_j\| \\ = \left(\sum_{k=1}^m (x_i^{(k)} - x_j^{(k)})^2 \right)^{1/2}, \quad i, j = 1, 2, \dots, N, \quad (1)$$

where $x_i^{(k)}$ denotes the k th component of x_i . Without loss of generality, we can assume that the x_i 's are centered.

Notice that the squared pairwise distance is

$$d_{ij}^2 = d^2(x_i, x_j) = \|x_i - x_j\|^2 = X_i^T x_i - 2x_i^t x_j + x_j^t x_i. \quad (2)$$

Let the N -dimensional vector ψ be

$$\psi = [x_1^T x_1, x_2^T x_2, \dots, x_N^T x_N]^T. \quad (3)$$

Then, the squared-distance matrix $D = [d_{ij}^2]_{N \times N}$ can be rewritten as

$$D = \psi e^T - 2X^T X + e \psi^T, \quad (4)$$

where e is the N -dimensional vector of all ones and $X = [x_1, x_2, \dots, x_N]$.

The mean-centered inner product matrix is defined as

$$H \equiv -\frac{JDJ}{2} = X^T X, \quad (5)$$

where $J = I - ee^T/N$ is the mean-centering matrix.

To recover X , let the eigendecomposition of H be

$$H = U \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N) U, \\ \lambda_1 \geq \dots \geq \lambda_p \geq \lambda_{p+1} \dots \geq \lambda_N, \quad (6)$$

$$U = [v_1, v_2, \dots, v_p, v_{p+1}, \dots, v_N],$$

where λ_i is the i th largest positive eigenvalue and v_i is the corresponding orthonormal eigenvector.

Then, the required p -dimensional embedding vectors Y are given by the columns of the following matrix:

$$Y = [Y_1, Y_2, \dots, Y_N] = \text{diag}[\lambda_1^{1/2}, \lambda_2^{1/2}, \dots, \lambda_p^{1/2}] U^{*T}, \quad (7)$$

where $U^* = [v_1, v_2, \dots, v_p]$.

MDS is widely used for the visualization of data and in molecular modeling. The popularity of MDS has also led to the proposal of its variants.

2.2. *Isomap*. When the data points lie on or close to a low-dimensional nonlinear manifold embedded in a high-dimensional space and the nonlinear structure cannot be adequately represented by a linear approximation, classical MDS as mentioned above usually fails to recover the low-dimensional structure of the nonlinear manifold. For example, the data points lying on an underlying manifold two-dimensional “Swiss-roll” embedded in a three-dimensional space cannot be reduced effectively by MDS method.

To deal with this problem, recently, a class of nonlinear embedding techniques has been designed to discover the structure of high-dimensional data and find their embedding in a low-dimensional Euclidean space. Among them, in 2000, Tenenbaum et al. [5] proposed a well-known dimensionality reduction technique, named Isomap, which combines the major algorithmic advantages of PCA and MDS. In the Isomap algorithm, geodesic distances between points are extracted instead of simply taking the Euclidean distance. The geodesic distances are computed by constructing a sparse graph in which each node is connected only to its closest neighbors. The geodesic distance between each pair of nodes is taken to be the length of the shortest path in the graph that connects them. These approximated geodesic distances are then used as input to classical MDS. Up to now, Isomap has proven to be successful in many applications, such as word inspection, visualization of biomedical data, and head pose estimation.

The Isomap algorithm can be written briefly as follows.

Step 1 (construct neighborhood graph). Determine which points are neighbors on the manifold M . Two simple methods are to connect each point to all points within some fixed radius ε (ε -Isomap) or to all of its k nearest neighbors (K -Isomap). These neighborhood relations are represented as a weighted graph G over the data points, with edges of weight $d_x(i, j)$ between neighboring points.

Step 2 (compute shortest paths). Isomap estimates the geodesic distances $d_M(i, j)$ between all pairs of points on the manifold M by computing their shortest path distance $d_G(i, j)$ in the graph G .

Step 3 (construct d -dimensional embedding). Apply classical MDS to the matrix of graph distance $D_G = d_G(i, j)$, constructing an embedding of the data in a low-dimensional Euclidean space Y that best preserves the estimated intrinsic geometry of the manifold.

K -Isomap implies that the KNN method is used in Step 1 in Isomap algorithm. The detailed procedure of Isomap algorithm can be referred to in [5].

3. An Improvement of K -Isomap

The basic idea of Isomap algorithm is that, for a neighborhood of points on a manifold, the Euclidean distances provide a fair approximation of geodesic distances. For faraway points the geodesic distance is estimated by the shortest path through neighboring points.

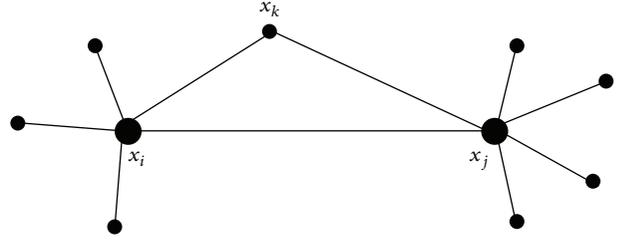


FIGURE 1: Five nearest neighbors of x_i and x_j .

It should be noted that the neighbors satisfy symmetrical relation in ε -Isomap algorithm. If data point x_j is in the neighbourhood of data point x_i , then x_i must be in the neighbourhood of x_j since the distance between x_i and x_j is less than or equal to ε . However, the symmetrical relation between neighbors in K -Isomap method is no longer right. As shown in Figure 1, if we take $k = 5$ nearest neighbors of x_i to construct neighborhood graph in K -Isomap, x_j is the fifth neighbor of x_i . But x_i is not one of the five nearest neighbors of x_j .

In this case, it is more reasonable to express the relation between x_i and x_j via x_k although the distance d_{ij} is less than $d_{ik} + d_{kj}$. To best preserve the close relation between neighbors properly in the procedure of dimensionality reduction, a novel technique is described to reduce high-dimensional data into a low-dimensional Euclidean space.

Suppose the data consist of N real-valued vectors, each dimensionality D , sampled from some underlying manifold. Provided there is sufficient data, we expect each data point and its neighbors to lie on or close to a locally linear patch of the manifold. We assign neighbors to each data point x_i based on the distances d_{ij} between pairs of points x_i and x_j by using the K nearest neighbors in the input space. To reflect the nearness of neighbor x_j to data point x_i , d'_{ij} is defined as

$$d'_{ij} = \frac{d_{ij}}{\sum_{n=1}^K d_{in}}, \quad (8)$$

where d_{in} denotes the distances between neighbor x_n and the data point x_i . If x_j does not belong to the set of neighbors of x_i , then $d_{ij} = \infty$.

It is obvious that d'_{ij} is generally not equal to d'_{ji} . In order to use MDS algorithm, therefore, we define the symmetrical matrix W as follows:

$$W = \left[\frac{(d'_{ij} + d'_{ji})}{2} \right]_{N \times N}. \quad (9)$$

The elements w_{ij} of matrix W indicate the average weights between x_i and x_j . The neighborhood relations are represented by a weighted graph G over the data points, with edges of weight w_{ij} between neighboring points. The graph G constructed by the proposed method is more effective to reflect the neighborhood relations between data points than that defined by K -Isomap. For instance, the edge connected data points x_i and x_j in Figure 1 do not exist according to

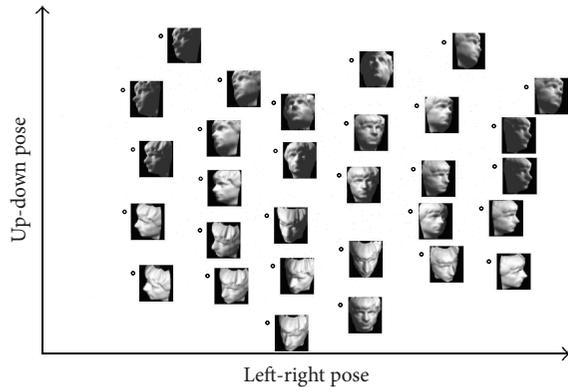


FIGURE 2: A two-dimensional projection of face dataset in 4096-dimensional space.

our method because the relation between them is not close together.

The geodesic distances w_{ij}^M between all pairs of points on the manifold M can be obtained by computing their shortest path distances w_{ij}^G in the graph G . As K -Isomap algorithm, we apply classical MDS to the matrix $D^G(w_{ij}^G)$, constructing an embedding of the data in a low-dimensional Euclidean space that best preserves the manifolds estimated intrinsic geometry.

4. Experiments

To verify the dimensionality reduction capability of the introduced method, we would like to test it on three classical examples which are widely used to test the effectiveness of dimensionality reduction techniques. The experimental platform is based on Windows 7 with AMD Phenom (tm) II P960 Quad-Core Processor 1.8 GHz and 2.00 GB memory. The programming language is Matlab R2011a.

In what follows, two classical pattern classification problems, face recognition and handwritten digit recognition, are considered in order to analyze the performance of our proposal.

One of the classical examples in dimensionality reduction from the domain of visual perception is face recognition. Multiple photographs ($N = 698$) of the same face with different poses and lighting directions are digitized as $64 \times 64 = 4096$ gray scale images. All of the images which are saved in no particular order can be considered as points in a high-dimensional vector space and lie on an intrinsically three-dimensional manifold. We test the improved method against this dataset (<http://isomap.stanford.edu/datasets.html>) and take $K = 8$. A two-dimensional projection is shown in Figure 2. The circles indicate the original input images and the points reflect the neighbor relations of the images. The direction of coordinate axis represents the gradual change of poses. The results show that the relations of images with similar poses are preserved perfectly on this dataset in the procedure of dimensionality reduction.

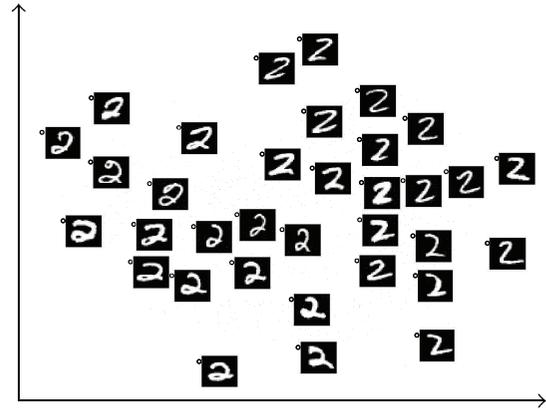


FIGURE 3: The neighborhood relations of 1032 the digits “2.”

The MNIST database of handwritten digits, available from <http://yann.lecun.com/exdb/mnist/>, has a training set of 60,000 examples and a test set of 10,000 examples. The digits have been size-normalized and centered in a fixed-size image. Each digit set in the dataset consists of about 1100 instances of a given numeral (“0” to “9”) scanned in as $28 \times 28 = 784$ pixel gray scale images. These handwritten digits vary in slope, angle, and thickness characters. In order to evaluate the performance of our proposal, $N = 1032$ the digits “2” which are written in various forms are selected from the MNIST database. The major features of the “2” are of bottom loop, top arch, and thickness. Applied to this dataset, the proposed method ($K = 19$) learns a two-dimensional embedding of the data’s intrinsic geometric structure. It is easy to see from Figure 3 that similar symbols are near each other. This fact indicates that the neighbors’ relation of data points in high-dimensional space still holds in two-dimensional space by using the introduced approach.

A low-dimensional nonlinear manifold embedded in a high-dimensional space is generally invisible, and therefore it is difficult to understand the dimensionality reduction clearly. To illustrate it effectively, we would like to introduce an artificial dataset “Swiss-roll” which is three-dimensional data ($N = 2000$) sampled from a two-dimensional manifold as shown in Figure 4(a). The colors of data points in Figure 4 indicate their neighborhood relations. It is easy to see from Figure 4(b) that the “Swiss-roll” is unfolded by the proposed approach ($K = 15$) in two-dimensional space. The example explains visually that the neighborhood relations of all data points in high-dimensional space are preserved well while reducing them into low-dimensional space by our proposal.

4.1. Discussions and Conclusions. Since two famous methods, Isomap and LLE, were proposed in 2000, the nonlinear techniques for dimensionality reduction have been widely investigated and successfully applied in various scientific fields, such as pattern recognition, visual perception of images of faces, handwriting, and gestures. Along this way, an improvement of K -Isomap algorithm is presented to reduce the dataset in high-dimensional space into a new dataset in low-dimensional space. The neighborhood relations can be

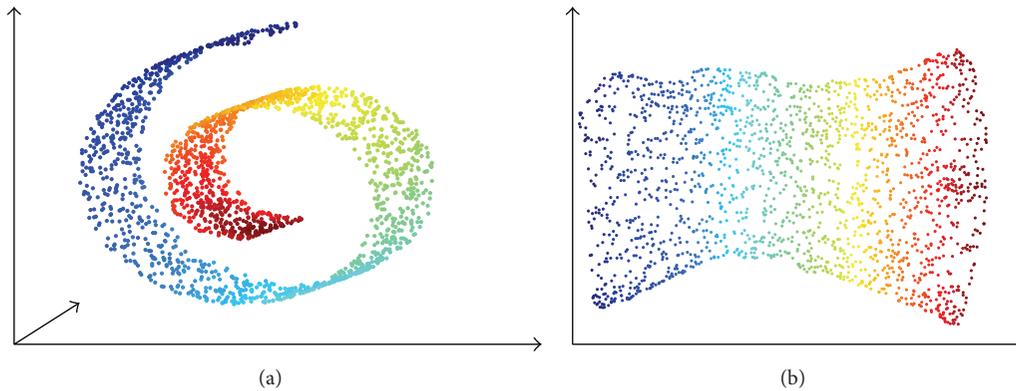


FIGURE 4: The “Swiss-roll” and its reduced result in two-dimensional space.

preserved well in the process of dimensionality reduction by the proposal. The reduced results of three classical examples account for the validity of our proposed method.

It is well known that the dimensionality reduction is a stage of data preprocessing in data mining, not our final destination. It is worthwhile to investigate how to apply the reduced results to classify/cluster the real/artificial datasets. Our future work will focus on this problem.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] L. van der Maaten, E. Postma, and J. van den Herik, “Dimensionality reduction: a comparative review,” *Journal of Machine Learning Research*, vol. 10, pp. 1–41, 2009.
- [2] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, USA, 2nd edition, 2002.
- [3] S. Mika, *Kernel fisher discriminant [Ph.D. thesis]*, University of Technology, Berlin, Germany, 2002.
- [4] W. S. Torgerson, “Multidimensional scaling: theory and method,” *Psychometrika*, vol. 17, pp. 401–419, 1952.
- [5] J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [6] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [7] H. Y. Zha and Z. Y. Zhang, “Isometric embedding and continuum ISOMAP,” in *Proceedings of the 20th International Conference on Machine Learning (ICML '03)*, pp. 864–871, Washington, DC, USA, 2003.
- [8] X. Q. Hu, Z. X. Yang, and L. Jing, “An incremental dimensionality reduction method on discriminant information for pattern classification,” *Pattern Recognition Letters*, vol. 30, pp. 1416–1423, 2009.
- [9] Y. L. Zheng, B. Fang, Y. Y. Tang, T. P. Zhang, and R. Z. Liu, “Learning orthogonal projections for Isomap,” *Neurocomputing*, vol. 103, pp. 149–154, 2013.
- [10] Y. D. Bu, F. Q. Chen, and J. C. Pan, “Stellar spectral subclasses classification based on Isomap and SVM,” *New Astronomy*, vol. 28, pp. 35–43, 2014.
- [11] C. Dimeglio, S. Gallón, J.-M. Loubes, and E. Maza, “A robust algorithm for template curve estimation based on manifold embedding,” *Computational Statistics & Data Analysis*, vol. 70, pp. 373–386, 2014.
- [12] H. Choi and S. Choi, “Robust kernel Isomap,” *Pattern Recognition*, vol. 40, pp. 853–862, 2007.
- [13] O. Samko, A. D. Marshall, and P. L. Rosin, “Selection of the optimal parameter value for the Isomap algorithm,” *Pattern Recognition Letters*, vol. 27, pp. 968–979, 2006.
- [14] B. Raducanu and F. Dornaika, “A supervised non-linear dimensionality reduction approach for manifold learning,” *Pattern Recognition*, vol. 45, pp. 2432–2444, 2012.
- [15] C. Williams, “On a connection between kernel PCA and metric multidimensional scaling,” *Machine Learning*, vol. 46, pp. 11–19, 2002.
- [16] T. Cox and M. Cox, *Multidimensional Scaling*, Chapman & Hall, London, UK, 1994.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

