Available online at www.sciencedirect.com
SCIENCE DIRECT
NEUROCOMPUTING
Neurocomputing 69 (2006) 1768–1771
www.elsevier.com/locate/neucom

Letters

ISOLLE: LLE with geodesic distance

Claudio Varini[a,b,*], Andreas Degenhard[b], Tim W. Nattkemper[a]

[a]Applied Neuroinformatics Group, Faculty of Technology, University of Bielefeld, Universitätsstrasse 25, 33615 Bielefeld, Germany
[b]Condensed Matter Theory Group, Department of Physics, University of Bielefeld, Universitätsstrasse 25, 33615 Bielefeld, Germany

Received 17 July 2005; received in revised form 12 December 2005; accepted 12 December 2005
Available online 23 February 2006
Communicated by R.W. Newcomb

Abstract

We propose an extension of the algorithm for nonlinear dimensional reduction locally linear embedding (LLE) based on the usage of the geodesic distance (ISOLLE). In LLE, each data point is reconstructed from a linear combination of its n nearest neighbors, which are typically found using the Euclidean distance. We show that the search for the neighbors performed with respect to the geodesic distance can lead to a more accurate preservation of the data structure. This is confirmed by experiments on both real-world and synthetic data.

Keywords: Geodesic distance; LLE; Nonlinear dimensional data reduction

1. Introduction

Algorithms for dimensional data reduction are particularly useful for discerning the information contained in high-dimensional data structures. Locally linear embedding (LLE) has recently been proposed as a method for the dimensional reduction of nonlinear high-dimensional data [4]. Its basic idea is intuitively to approximate nonlinear structures by an aggregation of small linear patches. A patch is given by a data point and its n nearest neighbors. The correlations between neighboring data points are mathematically expressed by a set of n weights which best characterize the local geometry within the patch. The high-dimensional data are then projected into a lower-dimensional space while preserving these neighboring correlations.

The number of neighbors n strongly influences the accuracy of the linear approximation of nonlinear data structures. Specifically, smaller the n, smaller the patch, and more faithful is the linear approximation. However, if these patches are disjoint, LLE can fail to detect the global data structure [3]. Disjoint regions can be obtained especially when the data points are spread among multiple clusters, or the data are sparse. The number of neighbors n therefore needs to be sufficiently high to prevent this failure. On the other hand, as the neighbors' search is typically conducted using the Euclidean distance, this can cause a data point to have neighbors which in fact are very distant as one considers the intrinsic geometry of the data. More intuitively, this can be imagined as a short circuit (see Fig. 1). The presence of short circuits is undesired, as they can cause LLE to misinterpret the real data structure.

To address the above outlined problems occurring to LLE when used with a relatively high n, we propose the usage of LLE with geodesic distance (ISOLLE). This metric has already been employed in other algorithms for nonlinear dimensional reductions such as isomap [5] and curvilinear distance analysis [2]. The geodesic distance between two data points can be intuitively thought of as their distance along the contour of an object (see Fig. 1). In this study, we show that the geodesic distance applied to the neighbors' search lowers the number of short circuits, thereby leading to a better preservation of the original data in the low-dimensional space. The efficacy of ISOLLE over LLE is tested on both synthetic and real-world data. In addition, a comparison of the running times is also performed.

*Corresponding author. Condensed Matter Theory Group, Department of Physics, University of Bielefeld, Universitätsstrasse 25, 33615 Bielefeld, Germany. Tel.: +49 521 106 6197; fax: +49 521 106 6455.
E-mail address: varini@physik.uni-bielefeld.de (C. Varini).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2005.12.120

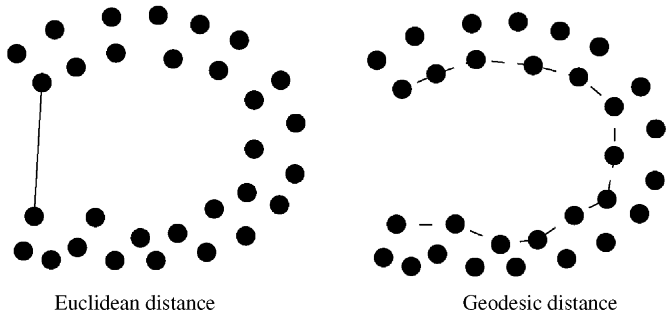

Fig. 1. The short circuit induced by Euclidean distance is shown on the left. According to this metric the two points appear deceptively close, although they are on the opposite parts of the horseshoe. On the right are shown the benefits of the geodesic distance. In this case the two points are not neighbors, as they are far away according to the geodesic distance.

2. Locally linear embedding (LLE)

Given V D-dimensional vectors $\{\mathbf{X}_i\}$ as input, the LLE algorithm comprises of three steps:

Step 1: it consists in assigning each data point $\mathbf{X}_i$ a predetermined number n of neighbors, typically according to the Euclidean distance;

Step 2: by minimizing the following error function

$$\Psi(W) = \sum_{i=1}^{V} \left| \mathbf{X}_i - \sum_{j=1}^{n} W_{ij}\mathbf{X}_j \right|^2 \tag{1}$$

with the condition $\sum_{j=1}^{n} W_{ij} = 1$, one computes the weights $\{W_{ij}\}$ that, combined with the neighbors, best approximate each data point $\mathbf{X}_i$;

Step 3: the weights are used to map each data point $\mathbf{X}_i$ to a d-dimensional vector $\mathbf{Y}_i$ with $d<D$, such that the error function

$$\Phi(Y) = \sum_{i=1}^{V} \left| \mathbf{Y}_i - \sum_{j=1}^{n} W_{ij}\mathbf{Y}_j \right|^2 \tag{2}$$

subject to $(1/V)\sum_{i=1}^{V} \mathbf{Y}_i\mathbf{Y}_i^{\mathrm{T}} = I$ and $\sum_{i=1}^{V} \mathbf{Y}_i = 0$, is minimized.

3. The ISOLLE algorithm

The ISOLLE algorithm differs from LLE in the first step, i.e. the neighbors' search. Specifically, ISOLLE computes the n nearest neighbors for each data point according to the geodesic distance. This process comprises two phases:

(a) define a graph G over all data points by connecting two points $\mathbf{X}_i$ and $\mathbf{X}_j$ if their Euclidean distance $d_{\mathrm{E}}(\mathbf{X}_i,\mathbf{X}_j)$ is lower than ε, and by assigning $d_{\mathrm{E}}(\mathbf{X}_i,\mathbf{X}_j)$ to the respective edge. The value of ε is chosen to be the minimum one which guarantees G to be connected;

(b) compute the n nearest neighbors for each data point according to the geodesic distances by Dijkstra's algorithm [1], which calculates the shortest path in G between any pair of points. Let $P(\mathbf{X}_1,\mathbf{X}_m) = \{\mathbf{X}_1,\ldots,\mathbf{X}_m\}$ be a path in G, i.e. $\mathbf{X}_i$ and $\mathbf{X}_{i+1}$ are connected for all $i \in \{1,2,\ldots,m-1\}$. Defining the path length

$$l(P(\mathbf{X}_1,\mathbf{X}_m)) = \sum_{i=1}^{m-1} d_{\mathrm{E}}(\mathbf{X}_i,\mathbf{X}_{i+1}), \tag{3}$$

the geodesic distance d_{G} between two points $\mathbf{X}_j$ and $\mathbf{X}_k$ is given by the length of the shortest path connecting them, that is

$$d_{\mathrm{G}}(\mathbf{X}_j,\mathbf{X}_k) = \min\{l(P(\mathbf{X}_j,\mathbf{X}_k))\} \quad \text{for all paths } P(\mathbf{X}_j,\mathbf{X}_k). \tag{4}$$

4. Data and experiments

We compare the performances of LLE and ISOLLE firstly on a synthetic data set, namely a three-dimensional swissroll (Fig. 2a), which is a convenient data set for illustrating the difference between LLE and ISOLLE and was already used in [4]; secondly on a complex, medical real-world data set acquired using dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) in breast cancer. DCE-MRI involves the repeated imaging of a region of interest after the administration of a contrast agent. In our case the DCE-MRI data of each patient comprise six volumes. Consequently, each voxel is associated with a time-series of six signal intensity values (Fig. 2b) which can be treated as a vector in $\Re^6$. The set of all time-series represents a multi-dimensional spatio-temporal data structure whose analysis can be eased by projecting the data into a lower-dimensional space. The data were provided by Dr. Axel Wismüller of the radiological department of the Munich University, Germany.

Both data sets are reduced to two dimensions using different values of n. By visualizing the neighbors' graphs obtained by connecting each data point in the swissroll with its n nearest neighbors, we can directly observe the lower number of short circuits induced by ISOLLE. Then, the benefits of the usage of the geodesic distance are illustrated by visualizing the respective two-dimensional projections of the swissroll.

In addition, the quality of the projections of both data sets is statistically evaluated in terms of topology preservation. Specifically, for each embedding we compute the *neighborhood preservation* (NP). NP is the average percentage of neighbors which are preserved after the

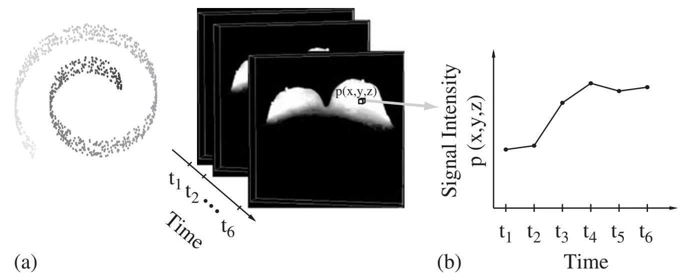

Fig. 2. (a) Three-dimensional swissroll data set. (b) In DCE-MRI, a time-series of MR signal intensity values is associated with each voxel $p(x,y,z)$.

dimensionality reduction. Mathematically, it is defined as

$$\mathrm{NP} = \frac{1}{V}\sum_{i=1}^{V} \mathrm{p}_t(\mathbf{X}_i), \tag{5}$$

where $\mathrm{p}_t(\mathbf{X}_i)$ is the percentage of the t nearest neighbors of point $\mathbf{X}_i$ in the original space which are preserved in the low-dimensional space. For example, if only 25% of its t nearest neighbors are preserved in the embedding, then $p_t(\mathbf{X}_i) = 0.25$. In this study we take $t = 5$. A high value of NP (close to 1) denotes a good preservation of the local relations between data points in the low-dimensional space and is therefore desirable.

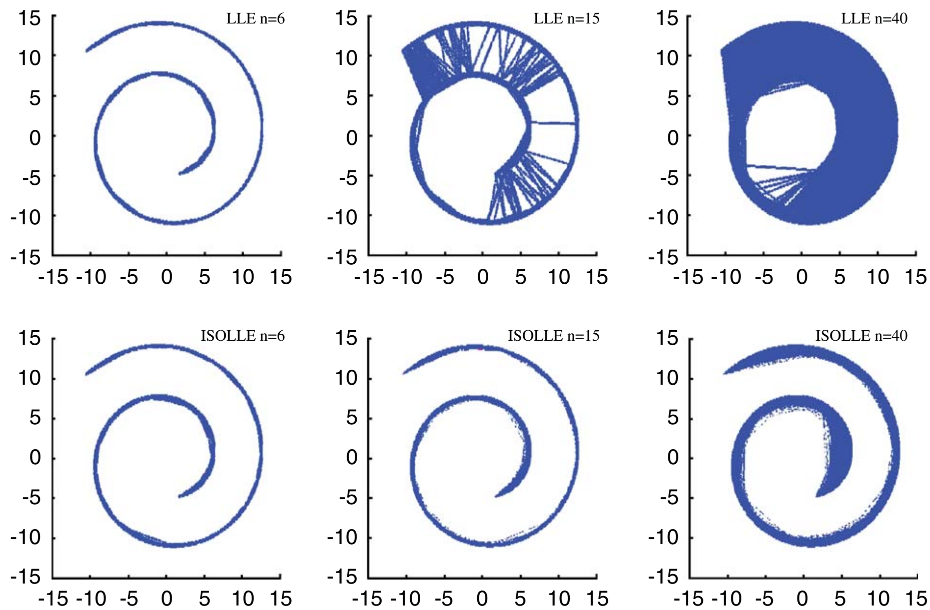

Fig. 3. Neighbors' graphs of the swissroll. In the one relative to LLE with $n = 15$ there are already short circuits. Their number considerably increases with $n = 40$. By contrast, in all the ISOLLE graphs there are no short circuits.

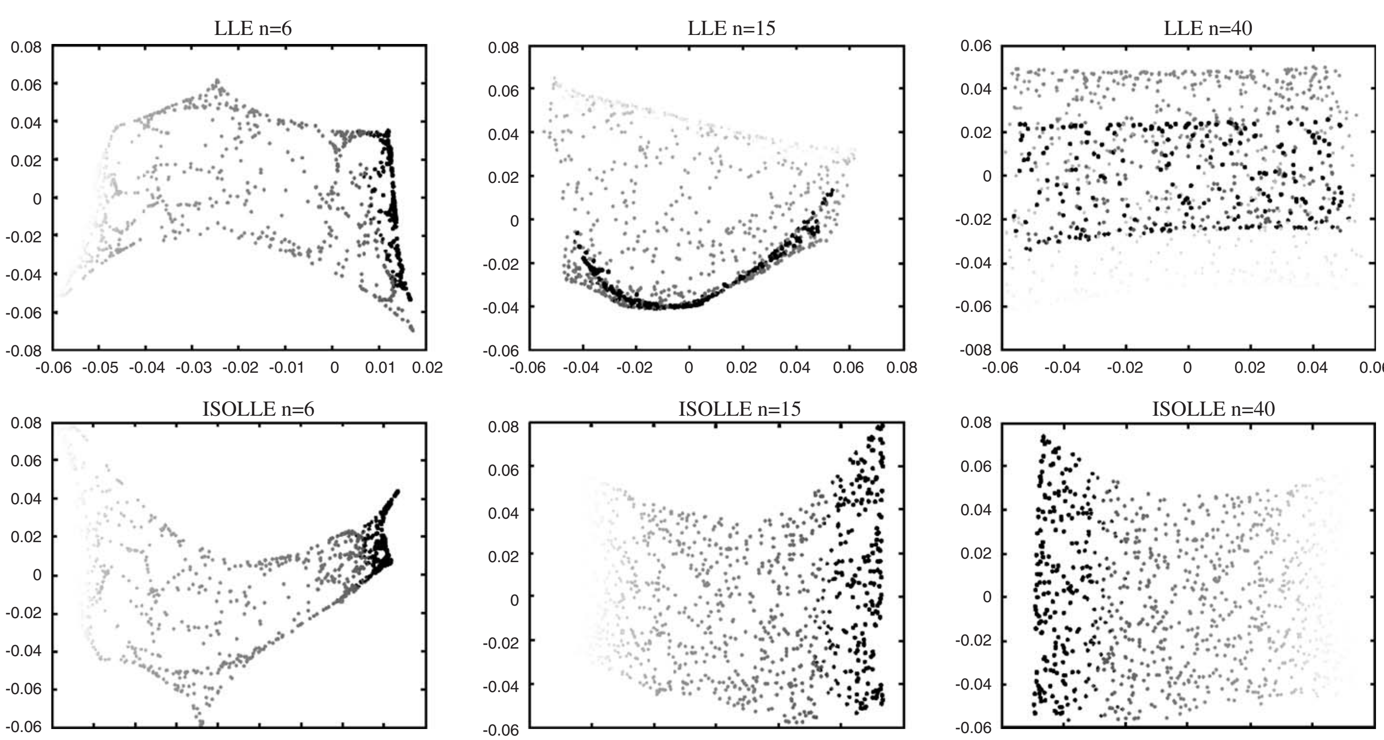

Fig. 4. Two-dimensional projections of the swissroll. LLE fails to unfold the swissroll with $N \geqslant 15$. Conversely, ISOLEE accurately projects the data in all cases.

5. Results and discussion

In Fig. 3, one can see the neighbors' graphs of the swissroll. It is obvious that, already with $n = 15$, LLE with Euclidean distance meets some short circuits in the neighbors' search. With $n = 40$, the number of short circuits increases considerably. By contrast, the graphs relative to ISOLLE do not present any short circuit, even with $n = 40$.

The effects of these short circuits can be observed in the two-dimensional projections of the swissroll in Fig. 4. Here, we see that LLE fails to preserve the global topological structure of the data with $n = 15$ and 40. Indeed, in both cases the darkest points are mapped closer to brighter points. On the contrary, ISOLLE can successfully unfold the swissroll in all three cases, and the projections are particularly good with $n = 15$ and 40.

We now compare the performances of LLE and ISOLLE in terms of NP. Its average value and variance, computed for the mappings relative to n comprised between 5 and 40, are displayed in Table 1. The low values of NP for the tumor data set reflect the complexity of the data. The values relative to ISOLLE are larger for both data sets, thereby suggesting that ISOLLE can better preserve the topology of the data.

Finally, we compare the running times of both algorithms and the values, obtained on a Pentium IV 2.8 GHz, are shown in Table 2. The usage of ISOLLE, as compared to LLE, involves a larger running time and this becomes more evident as n increases.

6. Conclusions

In this study, we show that the employment of LLE with geodesic distance (ISOLLE) can reduce the number of short circuits during the neighbors' search as compared to standard LLE with Euclidean distance, thereby allowing for a superior data preservation but at the cost of a larger running time. The efficacy of the ISOLLE algorithm is validated on both synthetic and real-world data. Future work will focus on the comparison between LLE and ISOLLE on more data sets, with particular regard to sparse data.

Table 1
Average and variance of neighborhood preservation (NP)

	Swissroll	DCE-MRI data
NP (LLE)	0.498 ± 0.004	0.081 ± 0.001
NP (ISOLLE)	0.737 ± 0.004	0.115 ± 0.002

Table 2
Table of the running times in seconds

	n	10	20	30	40
Swissroll ($V = 1000$, $D = 3$)	LLE	0.20	0.23	0.27	0.30
	ISOLLE	2.66	6.32	11.25	17.29
DCE-MRI data ($V = 2449$, $D = 6$)	LLE	1.26	1.39	1.62	1.75
	ISOLLE	16.55	39.24	69.31	106.37

References

[1] E.W. Dijkstra, A note on two problems in connection with graphs, Numerische Mathematik 1 (1959) 269–271.

[2] J.A. Lee, A. Lendasse, N. Donckers, M. Verleysen, A robust nonlinear projection method, in: Proceedings of European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium, April 2000, pp. 13–20.

[3] M. Polito, P. Perona, Grouping and dimensionality reduction by locally linear embedding, in: Advances in Neural Information Processing Systems (NIPS) 14, MIT Press, Vancouver, British Columbia, Canada, December 2001.

[4] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.

[5] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2322.

Claudio Varini received the Master Degree in Computer Engineering from the University of Modena, Italy in 2002. He is currently pursuing the Ph.D. degree in Computer Science in the Department of Physics and Computer Science at Bielefeld University, Bielefeld, Germany.

His research interests include machine learning, pattern recognition and explorative data analysis in biomedical domains.

Andreas Degenhard received his Master Degree in Theoretical Physics from the University of Muenster, Germany in 1995 and the Ph.D. degree in Mathematical Physics from the University of Bielefeld, Germany in 1999.

From 1999 to 2002 he worked as a postdoctoral researcher at the Institute of Cancer Research, London. Since 2002 he is employed as an assistant lecturer at the University of Bielefeld in the Theoretical Physics Department and employed consultant of the Institute of Cancer Research.

His fields of research include artificial neural network diagnosis, biomolecular adsorption and recognition processes, multiscale and multiresolution analysis, nonlinear partial differential equations and methods for complexity reduction in dynamical systems.

Tim Wilhelm Nattkemper received the Master degree and Ph.D. degree in Computer Science from the University of Bielefeld in 1998 and 2001, respectively. Since 2002 he is Junior professor at the Faculty of Technology, University of Bielefeld and head of the group "Applied Neuroinformatics".

In his research T.W.N. develops and evaluates new approaches for image analysis, multi-parameter data visualization and datamining in biomedicine which are based on machine learning algorithms. The application areas include segmentation and classification of microscopy images and tumors in multivariate magnetic resonance images. Other fields of research are text categorization, underwater image analysis and pattern recognition in proteomics and transcriptomics.