Available online at www.sciencedirect.com
SCIENCE DIRECT®

NEUROCOMPUTING

ELSEVIER

Neurocomputing 67 (2005) 29–53

www.elsevier.com/locate/neucom

Nonlinear dimensionality reduction of data manifolds with essential loops

John Aldo Lee*, Michel Verleysen[1]

Université Catholique de Louvain, Microelectronics Laboratory, Place du Levant 3, B-1348 Louvain-la-Neuve, Belgium

Received 29 February 2004; received in revised form 1 September 2004; accepted 18 November 2004

Communicated by S. Fiori

Abstract

Numerous methods or algorithms have been designed to solve the problem of nonlinear dimensionality reduction (NLDR). However, very few among them are able to embed efficiently 'circular' manifolds like cylinders or tori, which have one or more essential loops. This paper presents a simple and fast procedure that can tear or cut those manifolds, i.e. break their essential loops, in order to make their embedding in a low-dimensional space easier. The key idea is the following: starting from the available data points, the tearing procedure represents the underlying manifold by a graph and then builds a maximum subgraph with no loops anymore. Because it works with a graph, the procedure can preprocess data for all NLDR techniques that uses the same representation. Recent techniques using geodesic distances (Isomap, geodesic Sammon's mapping, geodesic CCA, etc.) or K-ary neighborhoods (LLE, hLLE, Laplacian eigenmaps) fall in that category. After describing the tearing procedure in details, the paper comments a few experimental results.

Keywords: Non-linear dimensionality reduction; Geodesic distance; Graph distance; Spanning tree; Graph cut

*Corresponding author. Tel.: +32 27 64 47 78; fax: +32 10 47 25 98.
E-mail addresses: lee@dice.ucl.ac.be (J.A. Lee), verleysen@dice.ucl.ac.be (M. Verleysen).
[1]M.V. works as a senior research associate of the Belgian FNRS.

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2004.11.042

1. Introduction

The standard task in nonlinear dimensionality reduction (NLDR, see [18]) consists in taking a set of data points sampled from a smooth, low-dimensional submanifold (e.g. P-dimensional) of a high-dimensional Euclidean space (e.g. D-dimensional), and to re-embed those points in a low-dimensional Euclidean space (e.g. P'-dimensional) while preserving the local structure of the submanifold. The purpose is to find useful new coordinates for the data. In general $D \gg P' \geqslant P$. If the original submanifold has the topology of a convex region in the P-dimensional Euclidean space, then one can hope for $P' = P$, and this is the best possible situation: the new coordinates parameterize the original submanifold. However, if the submanifold has a more complicated topology, perhaps having 'holes' or 'essential loops', then sometimes P' must be chosen greater than P because it is not possible or not as easy to find a P-dimensional embedding without tearing the manifold. For example, a torus is 2-dimensional but cannot be embedded in the 2-dimensional Euclidean plane.

This paper describes a procedure for 'tearing' or 'cutting' data manifolds, modifying them so that they have no essential loops anymore. Once this is done, in many cases one can then run NLDR algorithms successfully on the modified data, whereas the same algorithms might have failed or been less useful in the initial setting.

The proposed tearing procedure is actually applied to the 'neighborhood graph' which is often constructed in recent NLDR algorithms. Edges of the neighborhood graph connect data points which are close to each other in the high-dimensional Euclidean space. For example, each data point can be connected to its K-closest neighbors (K-ary neighborhoods) or to all other points lying no further than a certain distance ε from it (ε-neighborhood). NLDR algorithms using such a neighborhood graph are for instance LLE [15,17] and related techniques [7,2], or Isomap [21] and other algorithms using geodesic distances (e.g. geodesic versions of Sammon's nonlinear mapping (NLM) and curvilinear distance analysis [13]).

The structure of the paper is as follows. After this introduction, Section 2 briefly recalls some definitions about manifolds, graphs and shows how these two concepts are put together within the framework of NLDR. In particular, the last part of Section 2 explains how essential loops of manifolds are represented in their associated graph. Next, the tearing procedure itself is described in Section 3, whereas Section 4 shows a few experimental results on artificial data. Finally, Section 5 draws the conclusions and outlines perspectives for future work.

2. Manifolds, graphs and their relationship in NLDR

2.1. Manifolds and non-contractible loops

This section briefly defines some concepts about manifolds and graphs that are used further below.