Pattern Recognition Letters 38 (2014) 120–131

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Manifold learning by preserving distance orders ☆

Esra Ataer-Cansizoglu *,[1], Murat Akcakaya [1], Umut Orhan, Deniz Erdogmus

Cognitive Systems Laboratory, Northeastern University, Boston, MA, United States

ARTICLE INFO

Article history:
Received 13 May 2013
Available online 10 December 2013

Keywords:
Machine learning
Manifold learning
Nonlinear dimensionality reduction

ABSTRACT

Nonlinear dimensionality reduction is essential for the analysis and the interpretation of high dimensional data sets. In this manuscript, we propose a distance order preserving manifold learning algorithm that extends the basic mean-squared error cost function used mainly in multidimensional scaling (MDS)-based methods. We develop a constrained optimization problem by assuming explicit constraints on the order of distances in the low-dimensional space. In this optimization problem, as a generalization of MDS, instead of forcing a linear relationship between the distances in the high-dimensional original and low-dimensional projection space, we learn a non-decreasing relation approximated by radial basis functions. We compare the proposed method with existing manifold learning algorithms using synthetic datasets based on the commonly used residual variance and proposed percentage of violated distance orders metrics. We also perform experiments on a retinal image dataset used in Retinopathy of Prematurity (ROP) diagnosis.

1. Introduction

Due to the recent advances, acquisition of large volumes of high dimensional data has become more common in every aspect of daily life: stock market, social media, medical data, etc. Analysis and interpretation of such data requires finding meaningful low-dimensional structures in these huge data sets. Manifold learning attempts to accomplish such data explorations and dimensionality reductions.

Manifold learning can be regarded as identifying a nonlinear mapping from the original higher dimensional data space to a lower dimensional representation. Existing methods can be classified into three categories: global methods that tend to preserve global properties in the low-dimensional representation, local methods that aim to preserve the local geometry in the embedded space and techniques based on global alignment of multiple linear models (Van der Maaten et al., 2009). Multidimensional scaling (MDS) is one of the global methods that finds a projection of the original data while preserving the pairwise Euclidean distances (Kruskal, 1964). In the literature, various techniques are proposed to minimize MDS cost function (Dzwinel and Blasiak, 1999; Pawliczek et al., 2013; Pawliczek and Dzwinel, 2013; Andrecut, 2009). Similarly, in Isomap, one uses a geodesic distance estimation to use with MDS (Tenenbaum et al., 2000). Different variations of Isomap have been proposed in the literature: landmark and conformal Isomap (Silva and Tenenbaum, 2002). On the other hand, local methods (Zhang and Zha, 2004; Roweis and Saul, 2000; Weinberger and Saul, 2006; Dollár et al., 2006, 2007; Coifman and Lafon, 2006; Hinton and Roweis, 2002) constructs the lower dimensional data using the local linear relations in the original space. Local tangent space alignment (LTSA) (Zhang and Zha, 2004) represents the local geometry of the manifold with local tangent spaces that are learned through the neighborhood of each sample. Similarly, local linear embedding (LLE) (Roweis and Saul, 2000) aims to preserve local neighborhood information, while Semidefinite Embedding (SDE) (Weinberger and Saul, 2006) involves preserving local isometries on a k-nn graph. Coifman and Lafon (2006) presents a method that constructs local coordinates by learning a family of diffusion maps (DM). Another use of local geometry is by locally smooth manifold learning (LSML) (Dollár et al., 2006, 2007) which is based on learning a warping function, that takes any sample in the manifold and generates its neighbors. Stochastic neighborhood embedding (SNE) (Hinton and Roweis, 2002) and its variations (Xie et al., 2011; Van der Maaten and Hinton, 2008) are among probabilistic approaches that construct the neighborhood relations based on Gaussian kernels. Although local approaches have computational advantages, they might have limitations in preserving global geometry, especially if the data is sparse. Other methods that are based on global alignment of linear models aim to combine the local and global techniques by fitting a number of local linear models and merging them with a global alignment. Local linear coordination (LLC) (Teh and Roweis, 2002) and manifold charting (Brand, 2003) methods fall into this category.

☆ This paper has been recommended for acceptance by Eckart Michaelsen.

* Corresponding author. Tel.: +1 6173734779.

E-mail addresses: ataer@ece.neu.edu (E. Ataer-Cansizoglu), akcakaya@ece.neu.edu (M. Akcakaya), orhan@ece.neu.edu (U. Orhan), erdogmus@ece.neu.edu (D. Erdogmus).

[1] Esra Ataer-Cansizoglu and Murat Akcakaya contributed equally to this manuscript.

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.11.022

In this manuscript, we propose a nonlinear dimensionality reduction method that extends the basic idea used by the MDS and its variations. Although the ultimate goal is to preserve the distance orders, MDS algorithm only focuses on minimizing the mean-squared error between the input and output distances (Kruskal, 1964). There is no explicit constraint on the distance orders during the solution of this optimization problem. Moreover, this minimization results in a linear relationship between distance spaces. Linear fit assumption between the distances in the original and low-dimensional projection spaces is very restricted and embedding is achieved in this restrictive family. To address these two important issues, we first generalize the mean squared error cost function to include more general relations between distance spaces. Instead of assuming a predefined relationship, we propose to also learn the relationship between distances while we project the data from the original space. Our only assumption on the relationship is to have a monotonic nondecreasing function in order to preserve the distance relationships observed in the original space in the projected space. Then, we develop a constrained optimization problem by incorporating the distance orders as inequality constraints. As a solution of this problem, we not only learn the data in the projected space but also learn the non-linear relation between distance spaces. The final form of the proposed optimization problem is a generalization of the existing global MDS-based manifold learning algorithms such that the existing methods are approximate solutions of the simplified version of this problem. In this manuscript, we focus on the formulation and theoretical aspects of the problem. Possible acceleration of the proposed method by convex relaxations and further approximations to analyze real data will be part of our future research.

Another commonly used manifold learning algorithm which has a nonlinear mapping between distance spaces is Sammons mapping (Sammon, 1969). Sammon's map, a nonlinear extension of MDS, first maps the input data to a nonlinear predefined feature space and tries to preserve the distances in this feature space. Different than Sammon's mapping, we assume an unknown nonlinear relationship between the input and output distances while preserving the distance orders from the original space.

The rest of the paper is organized as follows: We first define the notation used throughout the paper in Section 2.1. Next, problem formulation is presented in Section 2.2. The solution of the optimization problem and performance evaluation metrics are explained in Sections 2.3 and 2.4 respectively. In Section 3, we report the experiments and results and the paper is concluded in Section 4.

2. Learning algorithm

In this section, we describe the proposed method for manifold learning. We first define the data model and notations to be used throughout the manuscript. Then using this model we formulate the desired manifold learning problem and develop our algorithm. We derive an optimization problem that solves the manifold learning algorithm, starting with the commonly used cost function, mean-squared error minimization, and demonstrate that this cost function can be extended to include different distance relations between the original and projection space data points, and explicit constraints that preserve distance orders in the projected space.

2.1. Data model and notation

We represent the original and the projected data spaces as $\mathcal{X}$ and $\mathcal{Y}$, respectively. Then, $\mathbf{x}_i \in \mathcal{X}$ and $\mathbf{y}_i \in \mathcal{Y}$ with $i = 1, \ldots, N$ are the data points. In this representation, $\mathbf{x}$ and $\mathbf{y}$ are vectors and N is the number of the data points. We assume that $\dim(\mathcal{X}) = d \geqslant \dim(\mathcal{Y}) = \tilde{d}$. Moreover, we have $d^x_{i,j}$ and $d^y_{i,j}$ as the distances between the ith and jth data points in the original and the low-dimensional data spaces, $||\cdot||$ represents the L2-norm of a vector.

2.2. Problem formulation

We formulate the manifold learning algorithm as a constrained optimization problem. Our approach restricts the minimum mean-squared error solutions used by some existing manifold learning algorithms (Kruskal, 1964; Tenenbaum et al., 2000). Specifically, these aim to minimize the difference between the distances of any two points in the original and projected spaces. That is, the difference between $d^y_{i,j}$ and $d^x_{i,j}$ for $\forall i,j = 1, \ldots, N$ is minimized, which on average results in a linear relationship between each $d^x_{i,j}$ and $d^y_{i,j}$ pair (as a result of the least-square solution).

$$\min_{\mathbf{y}_k \; k=1,\ldots,N} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} ||d^y_{i,j} - d^x_{i,j}||^2, \tag{1}$$

where $d^y_{i,j} = ||\mathbf{y}_i - \mathbf{y}_j||$ is the Euclidean distance and $d^x_{i,j}$ is the distance between the ith and jth points. Note here that the minimization is performed over the data points $\mathbf{y}_i$ in the projected space.

In the proposed algorithm, we compute $d^x_{i,j}$ as the estimated geodesic distance between the ith and jth data points. We follow the method described in Tenenbaum et al. (2000) to compute the geodesic distances in the original space. First, the Euclidean distance between every pair of data points in the original space (data pairwise distance matrix) is computed. Then, a k-nearest neighbor (knn) graph or ϵ-ball graph is generated. That is, k-nearest neighbors of a data point or neighbors within ϵ distance for each datum is taken, and the edge lengths from points outside these areas to the reference datum are set to be infinite, and the pairwise distance matrix is updated accordingly. Finally, Floyd algorithm is applied over this matrix to find approximate geodesic distances between the data pairs (Tenenbaum et al., 2000). Floyd's algorithm, an example of dynamic programming, finds the shortest path between each pair of vertices in a weighted graph (Floyd, 1962).

In our algorithm, we propose to generalize the minimum mean-squared error approach in (1) to include a broader relationship between the distances in the original and the low-dimensional projection space. For that purpose we rewrite (1) as

$$\min_{\mathbf{y}_k \; k=1,\ldots,N} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} ||d^y_{i,j} - h(d^x_{i,j})||^2, \tag{2}$$

where $h(\cdot)$ is a monotonic nondecreasing function. We represent the derivative of $h(\cdot)$ as

$$h'(d^x_{i,j}) = \sum_{k=1}^{N-1} \sum_{l=k+1}^{N} w_{kl} k(d^x_{i,j} - d^x_{k,l}), \tag{3}$$

where $k(\eta)$ is a translation-invariant kernel function and w_{kl}'s are the multiplicative coefficients. We force $w_{kl} \geqslant 0$ to have $h(\cdot)$ as a nondecreasing function. That is, we represent the derivative of $h(\cdot)$ as a nonnegative weighted sum of kernel functions. Monotonic nondecreasing functions $h(\cdot)$ will guarantee that we preserve the order of original distances in the projected space.

Then, we have

$$h(d^x_{i,j}) = \sum_{k=1}^{N-1} \sum_{l=k+1}^{N} w_{kl} K(d^x_{i,j} - d^x_{k,l}), \tag{4}$$

where