Pattern Recognition 47 (2014) 470–479

Contents lists available at ScienceDirect

Pattern Recognition

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

Inductive manifold learning using structured support vector machine

Kyoungok Kim [a], Daewon Lee [b,*]

[a] Department of Industrial and Management Engineering, POSTECH, San 31, Hyoja-Dong, Nam-Gu, Pohang, Kyungbuk 790-784, Republic of Korea
[b] School of Industrial Engineering, University of Ulsan, 93 Daehak-ro, Nam-gu, Ulsan 680-749, Republic of Korea

A R T I C L E I N F O

Article history:
Received 23 November 2012
Received in revised form
28 June 2013
Accepted 16 July 2013
Available online 29 July 2013

Keywords:
Dimensionality reduction
Manifold learning
Out-of-sample extension
Structured SVM

A B S T R A C T

Most manifold learning techniques are used to transform high-dimensional data sets into low-dimensional space. In the use of such techniques, after unseen data samples are added to the data set, retraining is usually necessary. However, retraining is a time-consuming process and no guarantee of the transformation into the exactly same coordinates, thus presenting a barrier to the application of manifold learning as a preprocessing step in predictive modeling. To solve this problem, learning a mapping from high-dimensional representations to low-dimensional coordinates is proposed via structured support vector machine. After training a mapping, low-dimensional representations of unobserved data samples can be easily predicted. Experiments on several datasets show that the proposed method outperforms the existing out-of-sample extension methods.

1. Introduction

Many examples of real-world high-dimensional data, such as images, are confined within a region of effective lower dimensional space, lying on a nonlinear sub-manifold embedded in some high-dimensional space. Manifold learning, the purpose of which is to find effective low-dimensional coordinates from high-dimensional data, has been intensively studied and applied to a variety of problems, such as computer vision, bioinformatics, econometrics, and signal processing [6,7,10,26].

Manifold learning is a part of dimensionality reduction, specifically nonlinear dimensionality reduction, with nonlinear methods including locally linear embedding (LLE) [33], Hessian LLE [5], local tangent space alignment (LTSA) [39], Kernel principal component analysis (PCA) [34], Isomap [35], autoencoder [14], self-organizing map (SOM) [19], maximum variance unfolding (MVU) [38] and so on. These nonlinear methods have some advantages over linear methods such as PCA [15] and linear discriminant analysis [9]. For instance, nonlinear methods can effectively capture nonlinear structures embedded in high-dimensional space. However, most such methods are transductive, which means retraining is necessary when unseen data samples are added after manifold learning techniques are applied to transform a high-dimensional data set into low-dimensional space.

This transductive property is the main limitation in the use of manifold learning as a preprocessing tool prior to the construction of predictive models in low-dimensional space. Dimensionality reduction techniques such as PCA are often used as a preprocessing step to avoid problems associated with dimensionality, reduce training and testing time, and improve the performance of the resulting model by removing redundant features. Here, unseen data samples are directly transformed into low-dimensional space to produce inputs of unseen data for the predictive model trained in the low-dimensional space. However, manifold learning methods generally do not yield a mapping from the original data (input space $\mathcal{X}$) to the reduced low-dimensional data (output space $\mathcal{Y}$). Instead, they should calculate low-dimensional coordinates by retraining training data with unseen data, whereas PCA and Kernel PCA compute low-dimensional representations using PC loading. In this case, mismatches between transformation results may occur, rendering the trained model useless. In addition, retraining requires a testing step with heavy time-complexity of $O(N^2)$ to compute relations between data points and $O(N^3)$ to solve the generalized eigenvalue problem for many manifold learning algorithms such as LLE and LTSA. Meanwhile several methods for out-of-sample extension have been proposed such as locally preserving projections (LPP) [13], neighborhood preserving embedding (NPE), and learning a eigenfunction [1]. However, because of the inherent linearity assumption of these methods, except autoencoder and SOM which train a mapping during the reduction process, they can not show good performance on nonlinear data. Besides, these methods also cannot provide a general framework to learn a mapping for any manifold learning.

In this paper, we propose a general framework for out-of-sample extension of *any* manifold learning. When given $\mathbf{x}\in\mathcal{X}$ and its corresponding reduced low dimensional data $\mathbf{y}\in\mathcal{Y}$ using *some* manifold learning, the proposed method learns a mapping from

* Corresponding author. Tel.: +82 52 259 1029; fax: +82 52 259 2180.
E-mail addresses: foriness@postech.ac.kr (K. Kim), dwlee0402@ulsan.ac.kr (D. Lee).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.07.011

$\mathcal{X}$ to $\mathcal{Y}$. For the learning task, we adapt the structured support vector machine (SVM) [2,23,36]. After training a mapping, obtaining low-dimensional coordinates well matched with those of training data becomes easy. The learning task can be viewed as a prediction of multivariate, continuous and unbounded output variables. Therefore, both of the conventional classification methods with discrete classes and regression methods for a univariate variable have limitations when it comes to solving this learning task.

The remainder of the paper is organized as follows: in Section 2, we briefly review the existing manifold learning methods and discuss their transductive property as a main limitation in their applications. The proposed inductive manifold learning method is presented in Section 3. In Section 4, we demonstrate the effectiveness of the proposed method using a synthetic and face data set, and apply the method to pedestrian and digit classification tasks. Finally, we conclude the paper in Section 5.

2. Transductive property of manifold learning

Dimensionality reduction is one of the main applications of manifold learning, also known as nonlinear dimensionality reduction. Nonlinear methods can be broadly classified into two groups: those that provide a mapping function and those that simply provide a representation. Autoencoder, SOM, and kernel PCA are examples of those methods that provide a mapping function, whereas most manifold learning techniques, such as LLE [33], Hessian LLE [5], LTSA [39], Isomap [35], MVU [38], diffusion map [21] and so on, only provide low-dimensional coordinates without mapping.

Dimensionality reduction is based on solving different optimization problems. In general, each method determines how to obtain proximity data, that is, distance measurements, and then establishes constraints and properties among data points in high-dimensional space which are also maintained in low-dimensional space. Their transductive property comes from the calculation of proximity data, because their distance relationships are changed when unobserved data points are added into the data set. Isomap, LLE, and LTSA are representative methods of transductive manifold learning and are briefly explained, including the underlying concepts and limitations as follows.

Instead of Euclidian distances, Isomap [35] calculates geodesic distances on a weighted graph. The multidimensional scaling method suffers from the fact that it does not take into account distribution [20], while Isomap captures relationships between data lying on or near a curved manifold, such as in the Swiss roll. In Isomap, geodesic distances between the data points $\mathbf{x}_i$ are computed by constructing neighborhood graph G in which every data point $\mathbf{x}_i$ is connected with its k nearest neighbors $\mathbf{x}_{ij}$ in the data set X. The low-dimensional representation $\mathbf{y}_i$ of the data point $\mathbf{x}_i$ in the low-dimensional space $\mathcal{Y}$ is computed by applying multidimensional scaling on the resulting distance matrix. Isomap must retrain data with new samples because the neighborhood graph G for computing geodesic distances between data points is updated after the addition of new samples.

LLE [5] is a nonlinear dimensionality reduction technique to preserve local relationships in a given data set, reducing sensitivity to short-circuiting. Consideration of local properties allows us to determine the embedding manifold of nonconvex structures. LLE describes the local relation around a D-dimensional data point $x_i \in \mathcal{D} \in \mathbb{R}^D$, $i = 1, \ldots, N$ on some manifold as a linear combination of its k nearest neighbors $\mathbf{x}_{ij} \in \mathcal{D}$ weighted by some w_{ij} which is obtained by minimizing reconstruction error, $\Phi(W) = \sum_i \|\mathbf{x}_i - \sum_j w_{ij}\mathbf{x}_{ij}\|^2$. Low-dimensional representations are calculated by minimizing reconstruction error with the same weight w_{ij} calculated in the high-dimensional space, $\Phi(Y) = \sum_i \|\mathbf{y}_i - \sum_j w_{ij}\mathbf{y}_{ij}\|^2 = YMY^T$, where the element of M is $m_{ij} = \delta_{ij} - w_{ij} - w_{ji} + \sum_k w_{ki} w_{kj}$. In LLE, low-dimensional representations are computed based on relationships among data samples; hence, all of relationships among data samples must be recalculated after the addition of unseen data samples to the data set.

LTSA [39] is a technique that describes local properties of high-dimensional data using the local tangent space of each data point. If local linearity of the manifold is assumed, there exists a linear mapping from a high-dimensional data point to its local tangent space and a linear mapping from the corresponding low-dimensional data point to the same local tangent space. LTSA simultaneously searches for the coordinates of low-dimensional data representations and linear mappings of low-dimensional data points to the local tangent space of high-dimensional data. Similar to LLE, LTSA calculates low-dimensional coordinates by solving a quadratic form of the optimization problem. Hence, LTSA must recalculate matrix M when unobserved samples are added to the data set, which means that LTSA is transductive.

As described above, most existing manifold learning methods depend on proximity expressions of training data without mapping, which makes a barrier to out-of-sample extension for them. Moreover, it is not a trivial matter to extend a variety of transductive proximity expressions into inductive ones. The better approach is therefore to directly learn a mapping from original space to its manifolds as a general framework for inductive manifold learning.

3. Inductive manifold learning

In this section, we describe the steps involved in learning a mapping for inductive manifold learning. Fig. 1 sketches the whole framework, which consists of learning a mapping and applying the learned mapping as a preprocessing step in predictive modelling such as classification or regression.

3.1. Modelling a mapping as structured SVM

Given a set of input points $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^D$ and their corresponding manifold points $\mathbf{y} \in \mathcal{Y} \subset \mathbb{R}^d$, we want to learn a mapping $f : \mathcal{X} \mapsto \mathcal{Y}$ with which we can predict dimensionality-reduced points on the manifold for unseen input $\mathbf{x}$ without retraining. We consider the case where the output space $\mathcal{Y} \subset \mathbb{R}^d$ is continuous, multivariate, and low dimensional manifold constructed by one of the plenty of manifold learning methods. We learn this mapping in the structured SVM framework [36] as

$$f(\mathbf{x}; w) = \underset{\mathbf{y} \in \mathcal{Y}}{\operatorname{argmax}}\, s(\mathbf{x}, \mathbf{y}; w), \tag{1}$$

where $s(\mathbf{x}, \mathbf{y}; w)$ is a discriminant function over input–output pairs that gives a large value to pairs $(\mathbf{x}, \mathbf{y})$ that are well matched and w represents a parameter vector. After learning the function s, we

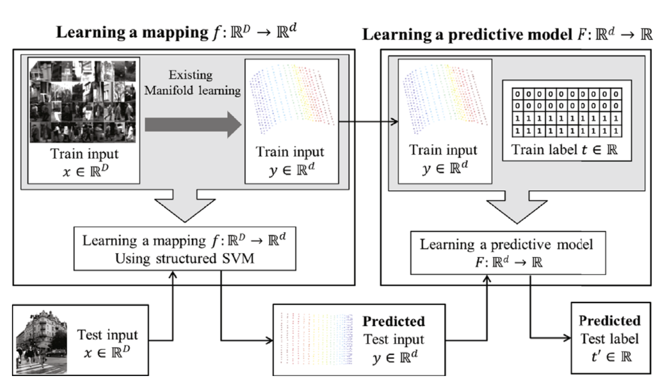

Fig. 1. A flowchart of the proposed inductive manifold learning framework to be used as a preprocessing step in a classification task.