

Kernel Isomap

Heeyoul Choi and Seungjin Choi ¹

Department of Computer Science
POSTECH
San 31 Hyoja-dong, Nam-gu
Pohang 790-784, Korea
{hychoi,seungjin}@postech.ac.kr

January 15, 2005

Abstract

Isomap [4] is a manifold learning algorithm, which extends classical multidimensional scaling (MDS) by considering approximate geodesic distance instead of Euclidean distance. The approximate geodesic distance matrix can be interpreted as a kernel matrix, which implies that Isomap can be solved by a kernel eigenvalue problem. However, the geodesic distance kernel matrix is not guaranteed to be positive semidefinite. In this letter we employ a *constant-adding* method, which leads to the Mercer kernel-based Isomap algorithm. Numerical experimental results with noisy "Swiss roll" data, confirm the validity and high performance of our *kernel Isomap* algorithm.

Indexing terms: Embedding, Kernel PCA, Manifold learning, MDS, Nonlinear dimensionality reduction.

Appeared in
Electronics Letters

vol. 40, no. 25, pp. 1612-1613, December 2004.

¹Please address correspondence to Prof. Seungjin Choi, Department of Computer Science, POSTECH, San 31 Hyoja-dong, Nam-gu, Pohang 790-784, Korea, Tel: +82-54-279-2259, Fax: +82-54-279-2299, Email: seungjin@postech.ac.kr

1 Introduction

Manifold learning involves inducing a smooth nonlinear low-dimensional manifold from a set of data points drawn from the manifold. Isomap is a representative isometric mapping, which extends metric MDS by considering approximate geodesic distance, instead of Euclidean distance [4].

Relationships between classical scaling and PCA were well studied [1]. The projection of the input data onto the eigenvectors of the data sample covariance matrix, returns the classical solution. If the kernel function is isotropic, then kernel PCA can be interpreted as performing metric MDS [5]. The approximate geodesic distance matrix computed in Isomap, can be interpreted as a kernel matrix [2]. However, the kernel matrix based on the doubly centered approximate geodesic distance matrix, is not always positive definite. We exploit a constant-adding method such that the geodesic distance-based kernel matrix is guaranteed to be positive semidefinite. Mercer kernel-based Isomap algorithm has a generalization property so that test data points can be successfully projected using a kernel trick as in kernel PCA [3], whereas in general embedding methods (including Isomap) do not have such a property.

2 The Kernel Isomap Algorithm

Given N objects with each object being represented by an m -dimensional vector $\mathbf{x}_t, t = 1, \dots, N$, the kernel Isomap algorithm finds a mapping which places N points in a low-dimensional space.

Algorithm Outline: Kernel Isomap

Step 1. Identify the k nearest neighbors of each input data point and construct a neighborhood graph where edge lengths between points in a neighborhood are set as their Euclidean distances.

Step 2. Compute approximate geodesic distances, d_{ij} , containing shortest paths between all pairs of points and define $\mathbf{D}^2 = [d_{ij}^2] \in \mathbb{R}^{N \times N}$.

Step 3. Construct a matrix $\mathbf{K}(\mathbf{D}^2)$ based on the approximate geodesic distance matrix \mathbf{D}^2 :

$$\mathbf{K}(\mathbf{D}^2) = -\frac{1}{2}\mathbf{H}\mathbf{D}^2\mathbf{H}, \tag{1}$$

where \mathbf{H} is the centering matrix, given by $\mathbf{H} = \mathbf{I} - \frac{1}{N}\mathbf{e}\mathbf{e}^T$, and $\mathbf{e} = [1, \dots, 1]^T \in \mathbb{R}^N$.

Step 4. Compute the largest eigenvalue, c^* , of the matrix

$$\begin{bmatrix} \mathbf{0} & 2\mathbf{K}(\mathbf{D}^2) \\ -\mathbf{I} & -4\mathbf{K}(\mathbf{D}) \end{bmatrix}, \quad (2)$$

and construct a Mercer kernel matrix $\widetilde{\mathbf{K}} = \widetilde{\mathbf{K}}(\widetilde{\mathbf{D}}^2)$ by

$$\widetilde{\mathbf{K}} = \mathbf{K}(\mathbf{D}^2) + 2c\mathbf{K}(\mathbf{D}) + \frac{1}{2}c^2\mathbf{H}, \quad (3)$$

where $\widetilde{\mathbf{K}}$ is guaranteed to be positive semidefinite for $c \geq c^*$.

Step 5. Compute the top n eigenvectors of $\widetilde{\mathbf{K}}$, which leads to the eigenvector matrix $\mathbf{V} \in \mathbb{R}^{N \times n}$ and the eigenvalue matrix $\mathbf{\Lambda} \in \mathbb{R}^{n \times n}$.

Step 6. The coordinates of the N points in the n -dimensional Euclidean space are given by $\mathbf{Y} = \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{V}^T$.

A main difference between the original Isomap and our kernel Isomap, lies in Step 4 which is related to the *additive constant problem* that was well studied in metric MDS. The additive constant problem aims at finding a value of constant, c , such that the dissimilarities defined by

$$\widetilde{d}_{ij} = d_{ij} + c(1 - \delta_{ij}) \quad (4)$$

have a Euclidean representation for all $c \geq c^*$ and δ_{ij} is the Kronecker delta. Substituting \widetilde{d}_{ij} for d_{ij} in (4) gives Eq. (3). For $\widetilde{\mathbf{K}}$ to be positive semidefinite, it is required that $\mathbf{x}^T\widetilde{\mathbf{K}}\mathbf{x} \geq 0$ for all \mathbf{x} . Cailliez showed that c^* is given by the largest eigenvalue of the matrix (2) (see Sec. 2.2.8 in [1]).

The matrix $\widetilde{\mathbf{K}}$ is a Mercer kernel matrix, so its (i, j) -element is represented by $\widetilde{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) = \phi^T(\mathbf{x}_i)\phi(\mathbf{x}_j)$ where $\phi(\cdot)$ is a nonlinear mapping onto a feature space or a low-dimensional manifold. The coordinates in the feature space can be easily computed by projecting the centered data matrix onto the normalized eigenvectors of the sample covariance matrix in the feature space, $\mathbf{C} = \frac{1}{N}(\mathbf{\Phi}\mathbf{H})(\mathbf{\Phi}\mathbf{H})^T$, where $\mathbf{\Phi} = [\phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_N)]$. The kernel Isomap algorithm provides a *generalization property (projection property)*, which provides a mapping for a test data point.

Given a test data point $\mathbf{x} \in \mathbb{R}^m$, its low-dimensional image, $\mathbf{y} \in \mathbb{R}^n$ is estimated by collecting projections onto the eigenvectors of \mathbf{C} . As in kernel PCA, the kernel trick leads to

$$y_j = \sum_{i=1}^N V_{ij} k(\mathbf{x}_i, \mathbf{x}), \quad (5)$$

where V_{ij} is the i th-element of \mathbf{v}_j and y_j is the j th-element of \mathbf{y} .

3 Numerical Experiments

We compare our kernel Isomap algorithm to the original Isomap algorithm, using Swiss roll data that was also used in Isomap. Noisy Swiss roll data was generated by adding isotropic Gaussian noise with zero mean and 0.5 variance (see Fig. 1 (a)). In the training phase, 1200 data points were used and the neighborhood graph was constructed using $k = 5$ nearest neighbors of each data point. As in Isomap, the shortest paths were computed using the Dijkstra’s algorithm, in order to calculate approximate geodesic distances. An exemplary embedding result (onto 3-dimensional feature space) for Isomap and kernel Isomap, is shown in Fig. 1 (b) and (c). Due to the space limitation, we show only this result, but for the case of noise-free Swiss roll data and noisy semi-sphere data, our kernel Isomap algorithm outperformed Isomap. The generalization property of our kernel Isomap is shown in Fig. 1 (d) where 3000 test data points are embedded with preserving local isometry well.

4 Conclusion

We have presented the kernel Isomap algorithm where the approximate geodesic distance matrix was interpreted as a kernel matrix and an adding-constant method was exploited so that the geodesic distance-based kernel became Mercer kernel. Main advantages of the kernel Isomap could be summarized as follows: (a) generalization property (i.e., test data points can be projected onto the feature space using the kernel trick as in kernel PCA); (b) robustness for noisy data. The generalization property will derive the kernel Isomap to be useful for pattern recognition problems.

5 Acknowledgment

This work was supported by Korea Ministry of Science and Technology under Brain Science and Engineering Research Program and International Cooperative Research Program and by KOSEF 2000-2-20500-009-5.

References

- [1] T. Cox and M. Cox. *Multidimensional Scaling*. Chapman & Hall, 2 edition, 2001.
- [2] J. Ham, D. D. Lee, S. Mika, and B. Schölkopf. A kernel view of the dimensionality reduction of manifolds. In *Proc. Int'l Conf. Machine Learning*, pages 369–376, Banff, Canada, 2004.
- [3] B. Schölkopf, A. J. Smola, and K. -R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, 1998.
- [4] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.
- [5] C. K. I. Williams. On a connection between kernel PCA and metric multidimensional scaling. *Machine Learning*, 46:11–19, 2002.

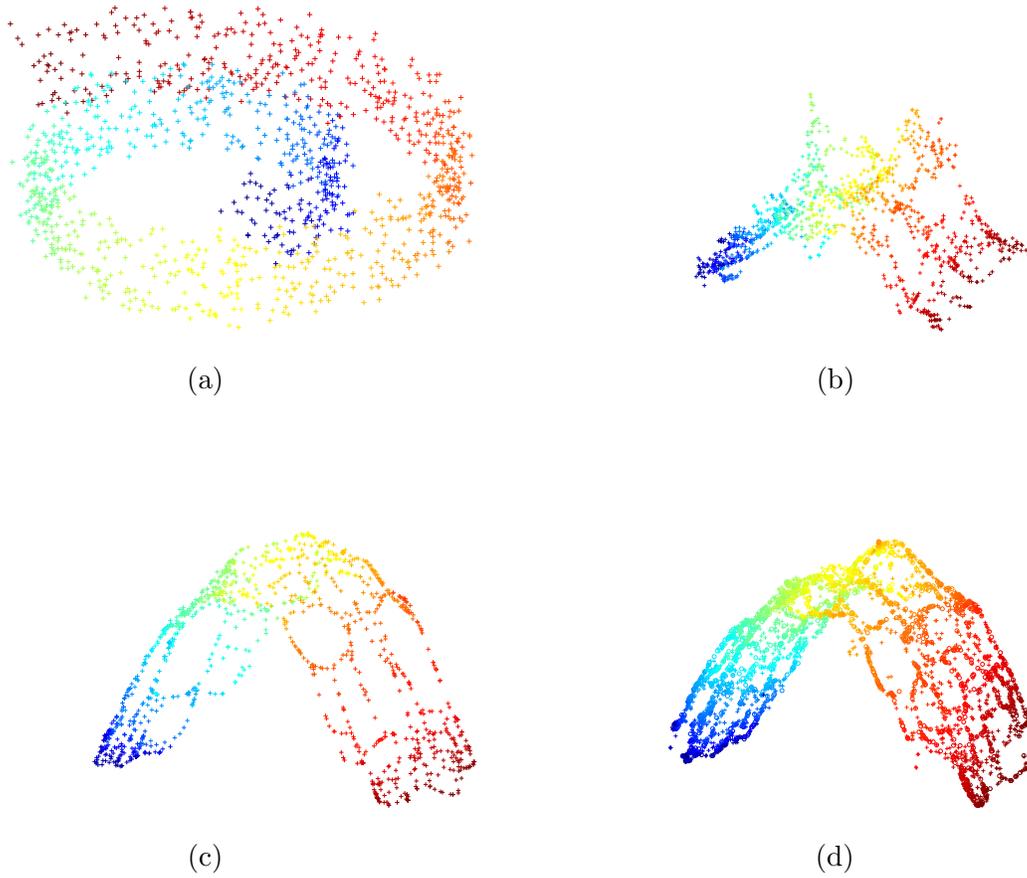


Figure 1: Comparison of the original Isomap with our kernel Isomap for the case of *noisy Swiss Roll* data: (a) noisy Swiss Roll data; (b) embedded points using the original Isomap; (c) embedded points using our kernel Isomap; (d) projection of test data points using the kernel Isomap. The modification by the constant-adding in the kernel Isomap improves the embedding with preserving local isometry (see (c)) as well as allowing to projecting test data points onto a feature space (see (d)).