Supervised Isomap with explicit mapping

CHUN-GUANG LI and JUN GUO
Beijing University of Posts and Telecommunications, Beijing, 100876, China
lcg444@126.com, guojun@bupt.edu.cn

Abstract

Isomap is one of the recently proposed manifold learning algorithms for nonlinear dimensionality reduction. However, Isomap not only suffers from a deficiency of no explicit mapping function, which is from high dimensional space to low dimensional space, but also does not employ the class information. In this paper, a supervised version of Isomap with explicit mapping, called SE-Isomap, is proposed. In SE-Isomap, geodesic distance matrix is calculated with respect to the class label information and Multidimensional Scaling (MDS) with explicit transformation is adopted instead of classical MDS used in Isomap. Thanks to the existence of explicit mapping and the use of class label information, SE-Isomap can be more easily used in pattern recognition than the original ones. Experimental results on two benchmark data sets demonstrated the performance of the presented method.

1. Introduction

Data in the real word are often complex to understand and difficult to process due to their high dimensionality. For this reason, one has to face the task to recover the meaningful low dimensional structure hidden in high dimensional observation data space. Manifold learning algorithms proposed in recent years are to cope with this task. Those manifold learning approaches include LLE [2], Isomap [3], Curvilinear Distance Analysis (CDA) [4], extended Triangulation [5] and some modified versions, such as C-Isomap and L-Isomap [6]. These methods attempt to best preserve the local neighborhood relationships of each point while utilize the rest points to keep the global information. All these methods, except CDA, suffer from the deficiency that lack of an explicit mapping, which can directly project an unseen sample into the target space. In addition class label information has not taken into account in all these methods. This paper focuses on to improve the original Isomap.

The original Isomap is not well suitable for multi-cluster data set and sensitive to noise. In the case of multi-cluster data set, [9] proposed a distance calculation strategy that the within-class geodesic distance adds the between-class Euclidean distance to solve the problem of disconnected neighborhood relationship graph. In [7], the linear weight function is used for decreasing the with-class distances. In [10] LDA step is adopted, instead of the classical MDS step in Isomap to make it optimal in the sense of classification. A so-called supervised Isomap (S-Isomap) is proposed in [8]. However, the S-Isomap could neither produce an explicit mapping nor deal with the multi-cluster data set.

In this paper, a version of Supervised Isomap with Explicit mapping, called SE-Isomap, is proposed by integrating the distance calculation strategy for multi-cluster data set and class label supervised information. Firstly, as in Isomap, the neighborhood relationship graphs for within-class data are constructed and the approximate geodesic distance matrix is calculated. Secondly the discriminative global distance matrix on the whole data set is constructed using the class label information and the strategy for multi-cluster data set. Finally MDS with explicit transformation [1] is utilized to obtain the low dimensional configuration and the explicit nonlinear mapping.

The remainder of this paper is organized as follows: in Section 2, the Supervised Isomap with Explicit mapping (SE-Isomap) will be presented. In Section 3, experiment results on two benchmark data sets will be shown. Finally, we will formulate our conclusions.

2. SE-Isomap algorithm

Given data set X which consisted of K classes $C_b (b = 1,...,K)$ that $X = C_1 \cup C_2 \cup ... \cup C_K$, where $C_b = \{x_i^b, x_i^b \in R^D, i = 1,...,n_b\}$, K is the number of classes and n_b is the number of samples within the

Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC'06)
0-7695-2616-0/06 $20.00 © 2006 IEEE

class C_b . For the reason of simplicity, let $K = 2$. Suppose that the intrinsic dimension of embedded manifold in R^D is d ($d \ll D$). SE-Isomap algorithm consists of three steps:

(1). Construct a neighborhood relationship graph $G_b(V, E)$ within each class $C_b (b = 1,2)$ and calculate the geodesic distance matrix G_{bb} respectively;

(2). Construct the discriminative global distance matrix G on the whole data set, $G = \begin{pmatrix} \rho_1 G_{11} & G_{12} \\ G_{21} & \rho_2 G_{22} \end{pmatrix}$ where G_{11} and G_{22} are the approximate geodesic distance matrices within class C_1 and C_2 respectively, G_{12} and G_{21} are the between-class distance matrix and $G_{12} = G_{21}{}^T$, $0 < \rho_1, \rho_2 \le 1$ are used for decreasing the within-class distance.

(3). Apply the MDS with explicit transformation on G to obtain both the low dimensional space configuration and the explicit mapping $F : R^D \to R^d$.

2.1. Calculation of geodesic distance matrix

Under the assumption that data points lie on a nonlinear manifold in high dimension space, so only for nearby points, Euclidean distance can be regarded as trustworthy measurement. Therefore, instead of Euclidean distance d_{ij} , the approximate geodesic distance $g_{ij} = g(i, j)$ is introduced to depict the intrinsic geometric structure of high dimension data set. For the within-class data, the geodesic distances are calculated using the shortest path on neighborhood relationship graph $G_b(V, E)$, as in Isomap.

2.2. Construct the discriminative global distance matrix

For the data points belong to different classes C_1 and C_2, the calculation of geodesic distance matrix G_{12} will take into account the strategy as [9]: for $i \in C_1$ and $j \in C_2$,

$$g(i, j) = g(i, i^*) + d(C_1, C_2) + g(j^*, j)$$

where $(i^*, j^*) = \underset{\forall p \in C_1, q \in C_2}{\arg\min}\, d\{p, q\}$, $\min_{\forall p \in C_1, q \in C_2} d\{p, q\}$ stand for the minimal between-class distance of classes C_1 and C_2 . $g(i, i^*)$ and $g(j^*, j)$ are the within-class geodesic distances. Notice that $d(C_1, C_2)$ is the between-class distance. For a visualization task, one would prefer to take $d(C_1, C_2) = \min_{\forall p \in C_1, q \in C_2} d\{p, q\}$ for trustworthiness. But then in the sense of classification, it may be optimal to take into consideration

$$d(C_1, C_2) = \lambda \max \left\{ \max_{i, j \in C_b} \{ d_G(i, j) \} \right\}_{b=1,2.}$$

for increasing the between-class separability, where $\lambda > 0$ is a constant for the tradeoff between the trustworthiness of visualization and the between-class separability.

On the other hand, ρ_1, ρ_2 are used to decrease the within-class pair-wise distances. In this way, the discriminative global distance matrix, which is beneficial to the classification task, is constructed.

2.3. MDS with explicit transformation

To obtain an explicit mapping of $R^D \to R^d$, MDS with explicit transformation [1] is applied instead of the classical MDS used in the original Isomap.

Let nonlinear mapping be $F : R^D \to R^d$, where $F = (f_1, ..., f_d)^T$. For $\forall x \in R^D$, we have

$$y = F(x) = W^T \Phi(x) \tag{1}$$

where W is $r \times d$ weight matrix, $\Phi(x) = (\phi_1(x), ..., \phi_r(x))^T$ is $r \times 1$ column vector, $\phi_j : R^D \to R (j = 1, ..., r)$ are a set of nonlinear basis functions, and $y \in R^d$ is d dimensional column vector in target space R^d . In fact, here we suppose that $f_i (i = 1, ..., d)$ can be written as a sum of fixed basis functions, that is for $x \in R^D$,

$$f_i(x|W) = \sum_{j=1}^{r} w_{ji} \phi_j(x)$$

where w_{ij} is the weight coefficient.

Inspired by the idea of Landmark-Isomap [6], some landmark points are also utilized to decrease computation complexity. The landmark points are selected by Locally Neighbor Covering algorithm (LNC) [12] within each class, respectively. Accordingly a new loss function is defined as:

$$\varepsilon(W) = \sum_{i=1}^{N} \sum_{j=1}^{L} \left(g_{ij} - q_{ij}(W) \right)^2$$

where Φ_i and Φ_j denote $\Phi(x_i)$ and $\Phi(x_j)$ respectively, $q_{ij}(W) = \left\| W^T (\Phi_i - \Phi_j) \right\|$ is the pair-wise Euclidean distance in target space R^d , and $L \ge d + 1$ is the number of landmark points to be selected. In this way, the problem turns to minimize $\varepsilon(W)$ with respect to W.

Iterative Majorization has been suggested for solving the least squares MDS problem [1], [11]. It can be summarized as follows:

1. Set $t=0$ and initialize $W^{(t)}$;
2. Set $V=W^{(t)}$;
3. Update $W^{(t)}$ to $W^{(t+1)}=\arg\min_{W}\varepsilon_m(W,V)$,

where $W^{(t+1)}$ is the solution of W as $W=A^{+}B(V)V$, where $A=\sum_{i,j}(\Phi_i-\Phi_j)(\Phi_i-\Phi_j)^T$, A^{+} is the Moore-Penrose inverse of A, and $B(V)=\sum_{i,j}b_{ij}(V)(\Phi_i-\Phi_j)(\Phi_i-\Phi_j)^T$ with

$$b_{ij}(V)=\begin{cases}\dfrac{g_{ij}}{q_{ij}(V)}, & q_{ij}(V)>0\\ 0 & , q_{ij}(V)=0\end{cases}.$$

4. Check for convergence. If not converged, then set $t=t+1$ go to step 2; otherwise stop.

2.4. Selection of parameters

The nonlinear basis functions need to be designated. Here we make choice of RBF for constructing the nonlinear mapping. The RBF can be expressed as:

$$\phi_i(x)=\exp\left\{-\frac{\|x-c_i\|^2}{\beta}\right\}$$

where $c_i(i=1,...,r)$ are a set of centers of $\phi_i(\cdot)$, r is the number of basis functions, and β is the bandwidth parameter. The expected centers $c_i(i=1,...,r)$ are those roughly represent the embedded manifold structure. In practice, those algorithms for data reduction can be adopted for the centers selection, such as Vector Quantization techniques. Here, the LNC [12] is used for RBF centers selection. The bandwidth parameter β will be determined according to the average distance of pair-wise RBF centers. Owning to the Iterative Majorization is not sensitive to the initial solution W^0, W^0 may utilize a random initialization.

Once the Iterative Majorization procedure converged, the optimal weight matrix W^* is found and the nonlinear mapping $F:R^D\to R^d$ is also obtained. For an unseen new sample $x_s\in R^D$, according to the formula (1), $y_s=F(x_s)=W^{*T}\Phi(x_s)$, where $y_s\in R^d$ is regarded as a feature vector for classification task.

3. Experimental results

In order to demonstrate the performance of our proposed method for supervised nonlinear dimensionality reduction, experiments on two benchmark data sets are shown here.

Firstly SE-Isomap is demonstrated on the classical Iris data set [13], which is consisted of 150 samples (50 of each of three classes). The 30 samples of each three class are used for training data and the rest 20 samples are treated as test data. As shown in Fig.1, (a) and (b) are the results of the original Isomap and the Extend Isomap (Ex-Isomap) [9] on the whole Iris data set where without the class label information is taken into account and k=8, β=10, r=45. (c) and (d) are the results of SE-Isomap on the training set and using the obtained mapping for test data respectively, where the class label information is adopted for increasing the between-class separability with $\lambda=0.8$, $\rho=0.9$. From (c) and (d), one can find that SE-Isomap make the between-class data more separable than Isomap (a).

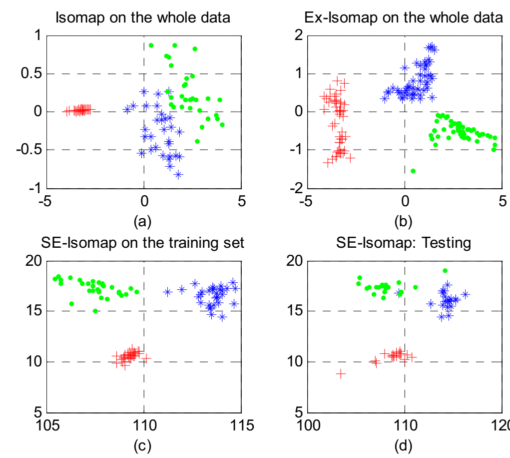

Figure 1. Experiment results on Iris data set

In the second experiments, 3 classes of COIL-20 images, totally 216 images (72 of each of 3 classes: Duck, Car and Cat) are used [14] and the results are shown in Fig. 2. Each image is resized into 32×32 (1024 dimensions) and the 60 images are used as training data for obtaining the nonlinear mapping and the rest 12 images are treated as testing data. (a) and (b) are the results of Isomap and Ex-Isomap [9], respectively (with k=17). (c) and (d) are the results of SE-Isomap on training data and using the obtained nonlinear mapping for testing data, where r=20, L=12 and k=3 for each class, $\lambda=0.5$, $\rho=1$, $\beta=5\times10^6$.

From (c) and (d), one can find that SE-Isomap make the between-class data more separable than Isomap (a), and manifold structure which lie in observation data approximately is preserved by SE-Isomap as in (c) and (d).

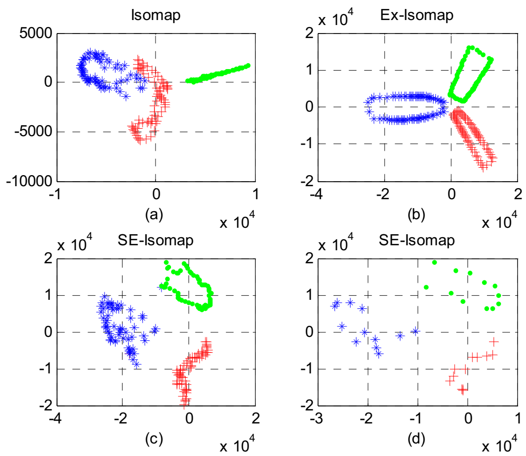

Figure 2. **Experiment results on COIL-20**

4. Conclusions and further works

In this paper, a supervised Isomap with explicit mapping, namely SE-Isomap, is proposed. In SE-Isomap, we adopted both the strategy for distance calculation on multi-cluster data set and class label supervised information to construct a discriminative global distance matrix. Moreover, the usage of MDS with explicit transformation, endue SE-Isomap with the ability to project an unseen data point, which is useful in pattern recognition.

However, Iterative Majorization suffers from the problem of not necessarily finding the global minimum. Therefore the approaches for searching a global minimum will be taken into account in further work.

Acknowledgements

This work was supported by National Natural Science Foundation of China under Grant No. 60475007.

References

[1] A. R. Webb, “Multidimensional scaling by Iterative Majorization using Radial Basis Functions,” Pattern Recognition, Vol. 28, No. 5, 1995, pp.753-759.

[2] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by Locally Linear Embedding,” Science, Vol.290, No.5500, Dec. 2000, pp.2323-2326.

[3] J. B. Tenenbaum, V. de Silva, and J.C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” Science, Vol.290, No.5500, Dec.2000, pp.2319-2323.

[4] J. A. Lee, A. Lendasse, N. Donckers, and M. Verleysen, “A robust nonlinear projection method,” Proc. ESANN 2000, Apr. 2000, pp.13-20.

[5]Y. Li, “Distance-preserving projection of high-dimensional data for nonlinear dimensionality reduction,” IEEE Trans. PAMI, Vol.26, No.9, Sept. 2004, pp.1243-1246.

[6] V. de Silva and J. B. Tenenbaum, “Global versus local methods in nonlinear dimensionality reduction,” in Advances in NIPS'2002, MIT Press, 2003, pp. 705-712.

[7] M. Vlachos, C. Domeniconi, D. Gunopulos, G. Kollios, and N. Koudas, “Non-linear dimensionality reduction techniques for classification and visualization,” Proc. 8th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, Edmonton, AB, Canada, pp.645-651, 2002.

[8] X. Geng, D. C. Zhang, and Z. H. Zhou, “Supervised nonlinear dimensionality reduction for visualization and classification,” IEEE Trans. on SMC-Part B: Cybernetics, Vol.35, No.6, 2005, pp.1098-1107.

[9] Y. M. Wu and K. L. Chan, “An extended Isomap algorithm for learning multi-class manifold,” Proc. ICMLC 2004, Vol.6, 2004, pp.3429-3433.

[10] M.-H. Yang, “Face recognition using extended Isomap,” Proc. ICIP, Sept 2002, pp.117-120.

[11] W. J. Heiser, “A generalized majorization method for least squares multidimensional scaling of pseudodistances that may be negative,” Psychometrika, Vol.56, No.1, 1991, pp.7-27.

[12] J. P. Zhang, “Manifold learning and applications,” PhD Thesis, Institute of Automation, Chinese Academy of Sciences, Beijing, China, 2003.

[13] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” Annual Eugenics, Vol.7, Part II, 1936, pp.179-188.

[14] S. A. Nene, S. K. Nayar and H. Murase, “Columbia Object Image Library (COIL-20),” Technical Report CUCS-005-96, February 1996.

0-7695-2616-0/06 $20.00 © 2006 IEEE