Proceedings of the Seventh International Conference on Machine Learning and Cybernetics,Kunming,12-15 July 2008

ADAPTIVE NEIGHBORHOOD SELECTION FOR MANIFOLD LEARNING

JIA WEI, HONG PENG, YI-SHEN LIN, ZHI-MAO HUANG, JIA-BING WANG

School of Computer Science and Engineering, South China University of Technology, Guangzhou 510641, China
E-MAIL: wei.jia@mail.scut.edu.cn

Abstract:

As a class of nonlinear dimensionality reduction methods, manifold learning can effectively construct nonlinear low dimensional manifolds from sampled data points embedded in high dimensional spaces. However, the results of most manifold learning algorithms are extremely sensitive to the parameters which control the selection of neighbors at each point. In this paper, an adaptive neighborhood selection method was proposed. Through ranking on manifold to select candidate neighborhood, and then estimating local tangent space, we can select the neighborhood of each point adaptively. Experimental results on several synthetic and real datasets demonstrate the effectiveness of our method.

Keywords:

Manifold Learning; Manifold ranking; Local tangent space; Adaptive neighborhood selection

1. Introduction

Manifold learning is a class of nonlinear dimensionality reduction techniques. It can effectively construct nonlinear low dimensional manifolds from sampled data points embedded in high dimensional spaces. The proposed algorithms include Isomap[1], locally linear embedding[2], Laplacian eigenmaps[3], Hessian eigenmaps[4], LTSA[5], and others. The first step of all the above algorithms is to construct the neighborhood of each point in the input space. It is critical to whether the algorithm will succeed or not. Two methods are commonly used for neighborhood selection: k nearest neighbors and neighbors within a fixed radius (ε hypersphere). The k nearest neighbors method is commonly used with respect to the sparseness of input space. There are many literature about the problem of selecting the parameter k. Samko et al[6] made use of residual variance to find the global optimal k, and Wen et al[7] used graph algebra to optimize neighborhood selection. There are several shortcomings of these methods. Firstly, they are time-consuming. Secondly, when the curvature and density of input space vary very much with the manifold, there may be no global optimal setting of k which can yields satisfying results. To solving this problem, some adaptive neighborhood selection methods were proposed. Wang et al[8] proposed an adaptive neighborhood selection method based on estimation of local tangent space, and apply it to LTSA. It includes two major steps of neighborhood contraction and neighborhood expansion, but the neighborhood size is indirectly controlled by several user-specified parameters, and it may result in the problems of 'short circuit' and 'smooth out' in the manifold[9]. Mekuz et al[9] proposed a heuristic method for selecting a neighborhood size adaptively that does not require any parameters. The method is based on estimation of intrinsic dimensionality and tangent space, but it relies on the density of sampling very much, when the input data is sparsely sampled it may fail to yield satisfying results.

In this paper, we describe a practical strategy for selecting adaptive neighborhood. The method combines the virtue of the above methods. It can select neighborhood adaptively and avoid the 'shortcuts' and 'smooth out' and isn't sensitive to density of sampling. The key idea of the method is to rank the neighborhood points using manifold ranking method[10] to select candidate neighbors of each point and then construct suitable local tangent space using heuristic method and select the optimal k from the candidate neighbors. In the experiments, we apply our method to LTSA and Isomap and demonstrate that our method can generate satisfactory results, while other methods fail.

2. Ranking on manifold and its relationship with neighborhood selection

At the first step of the manifold learning algorithms, we usually select the neighborhood of an input point using k nearest neighbors method that is selecting neighbors through ranking by Euclidean distance. But in many cases, this is not always correct. As Figure 1 demonstrates, the 4 nearest neighbors of point 'a' according to the ranking by Euclidean distance is {c, b, g, f}. Obviously, the short circuit problem occurs here. By contrast, if ranking on manifold, the 4 nearest neighbors of point 'a' should be {c, b, d, e}. This is a more proper neighborhood which can

978-1-4244-2096-4/08/$25.00 ©2008 IEEE

indicate its neighbors on the manifold and can ensure to preserve the intrinsic structure of the data set.

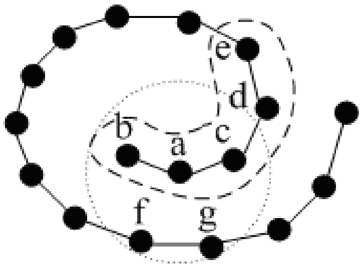


Figure 1. The different neighborhoods using ranking by Euclidean distance and ranking on manifold.

The core idea of ranking on manifold is to rank the data with respect to the intrinsic manifold structure collectively revealed by a great amount of sampled input data, and can be viewed as an case of semi-supervised learning[10].

Given a set of points $X = \{x_1,...,x_q, x_{q+1},...,x_n\} \subset R^m$, the first q points are the queries and the rest are the points that we want to rank according to their relevances to the queries. Let $d : X \times X \to R$ denote a metric on X , such as Euclidean distance, which assigns each pair of points x_i and x_j a distance $d(x_i, x_j)$. Let $f : X \to R$ denote a ranking function which assigns each point x_i a ranking value f_i . So f can be viewed as a vector $f = [f_1,...f_n]^T$. We also define a vector $y = [y_1,...y_n]^T$, in which $y_i = 1$ if x_i is a query, and $y_i = 0$ otherwise.

The ranking algorithm is as follows[10]:

1) Sort the pairwise distances among points in ascending order. Repeat connecting the two points with an edge according the order until a connected graph is obtained.

2) Form the affinity matrix W defined by $W_{ij} = \exp[-d^2(x_i, x_j)/2\sigma^2]$ if there is an edge linking x_i and x_j . Because there are no loops in the graph, we set $W_{ii} = 0$. In our experiments, we found that the algorithm is not sensitive to the value of σ , and we set $\sigma = 1$ in the following experiments and can get satisfying results.

3) Symmetrically normalize W by $S = D^{-1/2}WD^{-1/2}$ in which D is the diagonal matrix with D_{ii} equal to the sum of the i -th row of W .

4) Iterate $f(t+1) = \alpha Sf(t) + (1-\alpha)y$ until converges at a certain value, where α is a parameter in $[0,1)$.

5) Let f_i^* denote the limit of the sequence $\{f_i(t)\}$. Rank each point x_i according its ranking score f_i^* (largest ranked first).

About the convergence of this algorithm, it has the following theorem[11]: The sequence $\{f(t)\}$ converges to

$$f^* = (1-\alpha)(I - \alpha S)^{-1} y \qquad (1)$$

where I is the unit matrix.

We can now use the algorithm above to select a point's candidate neighbors. For each point $x_i \in X$, we set $y_i = 1$ and set other elements of vector y to 0. In the following experiments of this paper, we set $\alpha = 0.99$. Then using equation 1, we can get the ranking scores of all the points, and select the points which have the k largest ranking scores as the candidate neighbors of x_i .

Without special remark, the following ‘ k nearest neighbors’ all means the points which have the k largest ranking scores on the manifold, not the points ranking by Euclidean distance.

3. Estimating Local Tangent Space

In this section we will introduce a heuristic method to estimate local tangent space of each input point. Given a set of points $X = \{x_1,...,x_n\} \subset R^m$ whos intrinsic dimensionality is d , for each point $x_i \in X$, $X_i = \{x_{i1},...,x_{ik}\}$ is the k nearest neighbors of point x_i , then it is proved that

$$r_i^{(k)} = \sqrt{\frac{\sum_{j>d} (\sigma_j^{(k,i)})^2}{\sum_{j\le d} (\sigma_j^{(k,i)})^2}} \qquad (2)$$

can be viewed as the relative error between X_i and its local tangent space[8][12], where $\sigma_1^{(k,i)} \ge ... \ge \sigma_d^{(k,i)} \ge \sigma_{d+1}^{(k,i)} \ge ... \ge \sigma_k^{(k,i)}$ are the singular values of matrix $X_i - x_i 1_k^T$. When $r_i^{(k)}$ is close to zero, it means that the local tangent space can be reliably estimated form X_i . So, the problem is to find the suitable k so as to we can estimate the accurate local tangent space of x_i from X_i .

In order to ensure a sound d dimensional basis of local tangent space of point x_i , $\sigma_d^{(k,i)}$ should be sufficiently large and $\sigma_{d+1}^{(k,i)}$ should be sufficiently small. So we present a heuristic method to find the suitable k which can satisfy the above condition: iteratively increase k ($k \ge d+1$), and generate a new relative error $r_i^{(k)}$ in each iteration until

$$r_i^{(k)} \geq (\sigma_{d+1}^{(k,i)})^2 / (\sigma_d^{(k,i)})^2 \qquad (3)$$

and then the orthogonal basis of local tangent space at point x_i can be denoted as $A_i \in R^{m \times d}$, where the columns of A_i are the eigenvectors corresponding to the d largest singular values of $X_i - x_i 1_k^T$.

4. Adaptive Neighborhood Selection

Using the orthogonal basis A_i which denotes the estimated local tangent space at point x_i, we can select its neighborhood adaptively. As Figure 2 demonstrated, the line from 'a' to 'c' represent the local tangent space at point 'a', and 'c' is the projection of 'b' onto the local tangent space. It is easily to known that the length between 'a' and 'c' is $\left|A_a^T(x_b - x_a)\right|_2$, and the length between 'a' and 'b' is $\left|(x_b - x_a)\right|_2$, so we can use the following ratio

$$delta_{ba} = \left|A_a^T(x_b - x_a)\right|_2 \Big/ \left|(x_b - x_a)\right|_2 \qquad (4)$$

to determine whether 'b' is the neighbor of 'a' or not. If $delta_{ba} > \eta$, then 'b' can be considered as the neighbor of 'a', where η is a free parameter.

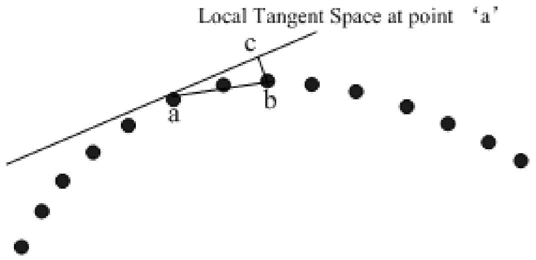


Figure 2. Schematic plot of adaptive neighborhood selection.

It is found that we can get satisfying results if $\eta \in [0.85, 0.95]$ in the synthetic experiments. We incrementally grow the neighborhood of 'a' according to the descending order of ranking scores of its k nearest neighbors one point a time, monitoring each new point's projection on to the local tangent space at 'a', and calculate its distance to 'a'. New neighbor x_b is added iteratively until $delta_{ba} < \eta$ and the iteration terminates.

With the above preparation, we can now present the adaptive neighborhood selection algorithm. Given a data set $X = \{x_1, ..., x_n\} \subset R^m$ whos intrinsic dimensionality is d, the approach consists of the following steps:

Step 1. Ranking on manifold and selecting candidate neighborhood: for each point x_i, selecting its k nearest neighbors according to the ranking scores on the manifold using the method introduced in Section 2. In this step, we set $k = k_{\max}$, where $k_{\max}$ is the largest value of k in the equation $\frac{2 \times P}{N} \leq k + 2$, here P is the number of edges and N is the number of nodes in the neighborhood graph from the first step of ranking algorithm[6].

Step 2. Estimating local tangent space: for each point x_i, using the method proposed in Section 3 to estimate its local tangent space according to the neighbors selected in the previous step.

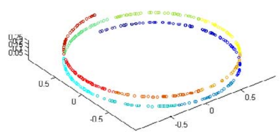
Step 3. Adaptive neighborhood selection: for each point x_i, selecting its neighbors according to the ratio *delta* of its candidate neighbors using the method proposed in Section 4. If samples were density sampled, we set *delta* large, or small otherwise.

5. Experimental Results and Analyses

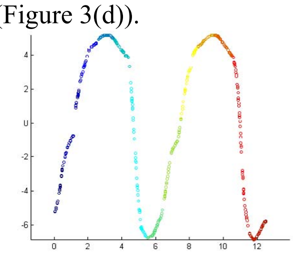
In this section we have tested our method with LTSA algorithm and Isomap algorithm on several synthetic and real datasets to show the effectivity and advantage of it.

5.1. Compressed Helix Dataset

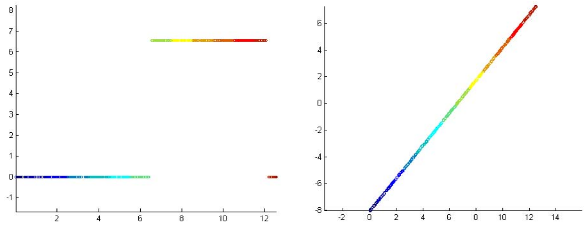
We sampled 500 data points from a compressed helix $x_i = [\sin(t_i); \cos(t_i); 0.02 * t_i]^T$, where $i = 1, ..., 500$, and $t_i \in [0, 4\pi]$ are uniformly distributed, and then use LTSA algorithm to get the low dimensional embedding. It is known that the arc length of a point x_i is $\sqrt{1 + 0.02^2} * t_i$, and it can be easily known that using arc length as the x-axis and low dimensional embedding as the y-axis we can get a slope line. As Figure 3 demonstrated, only our method can generate satisfying result(Figure 3(d)).



(a)



(b)

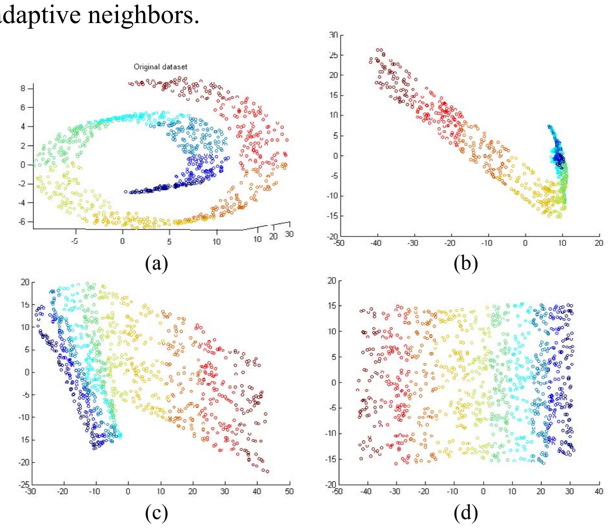


(c) (d)

Figure 3. (a).Compressed helix dataset($n=500$). (b).Embedding computed by original LTSA($k=10$). (c). Result of Mekuz et al's method using LTSA. (d). The projection obtained by LTSA using our adaptive neighborhood selection method($\sigma=1$, $\eta=0.95$).

5.2. Stretched Swiss Roll Dataset

In this experiment, we apply different neighborhood selection methods to the stretched swiss roll data set, and use Isomap algorithm to get the low dimensional embedding. We sampled 1000 points from the stretched Swiss roll dataset: $x_i=[t_i*\cos(t_i);\ heigh_i;0.6*t_i*\sin(t_i)]^T$ $+0.05*rand(1,3)$,where $i=1,...,1000$,and $t_i\in[3\pi/2$ $,9\pi/2]$ are uniformly distributed, and $heigh_i\in[0,30]$ are uniformly distributed too, as Figure 4(a) demonstrated. It can be seen that Figure 4(b) using Euclidean distance for neighborhood selection get a bad result. Figure 4(c) get a better result which using Mekuz et al's method. The result of Figure 4(d) is the best which using our method to select adaptive neighbors.



(c) (d)

Figure 4. (a).Stretched swiss roll dataset($n=1000$). (b).Embedding computed by original Isomap($k=10$). (c). Result of Mekuz et al's method using Isomap. (d). The projection obtained by Isomap using our adaptive neighborhood selection method($\sigma=1$, $\eta=0.9$).

We can use residual variance to evaluate the embedding result of different methods using Isomap[1]. Residual variance is defined as $1-r^2_{\hat{D}_X D_Y}$, where $\hat{D}_X$ are the matrices of geodesic distances between pairs of points in input space and D_Y are the matrices of Euclidean distances between pairs of points in output space, $r_{\hat{D}_X D_Y}$ is the standard linear correlation coefficient, taken over all entries of $\hat{D}_X$ and D_Y . So the smaller the residual variance is the better the embedding result is. The residual variances of the three neighborhood selection method are showed in table 1. It can be seen that the residual variance of our method is the smallest of the three. So it demonstrates that the neighborhood selection of our method is the most suitable one.

Table 1. The Residual Variance of Different Neighborhood Selection Methods.

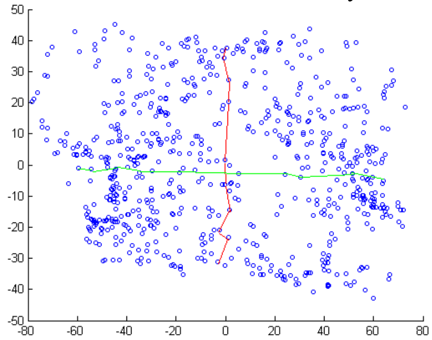
	Residual variance
Original Isomap ($k=10$)	0.6558
Mekuz et al's method	0.2697
Our method	0.0013

5.3. Sculpture Face Dataset

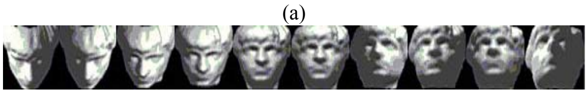
In this experiment, we will demonstrate that our method is also useful to real dataset. The Sculpture face dataset consists of 698 64-by-64 pixel images of sculpture faces under varying pose and illumination[1]. Each image can thought to be a $m=4096$ dimensional image vector in high dimensional space. Some of the images are showed in Figure 5.

Figure 5. Some examples of sculpture face dataset.

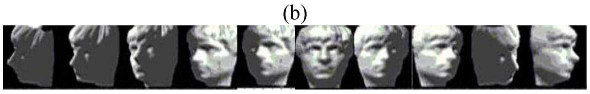
To reduce computational time, we first use PCA[13][14] method to the original data (99 percent of the principal component kept), and then apply our method to find adaptive neighbors ($\sigma = 1, \eta = 0.4$) and then use Isomap algorithm to construct low dimensional embedding as Figure 6(a) demonstrate. We also extracted two paths along the middle of the set of the 2D coordinates, and display the corresponding images along each path. It can be seen that the computed 2D coordinates do capture the pose and illumination variations in a continuous way.



(a)



(b)



(c)

Figure 6: (a).Embedding results computed by Isomap with adaptive neighborhood selection. (b). Images corresponding to red line(from top to bottom). (c). Images corresponding to green line(from left to right).

6. Conclusion

It is known that in many situations there doesn't exist an optimal global setting of neighborhood k for manifold learning. So this paper proposed an effective way of adaptive neighborhood selection at each point on a nonlinear manifold. Our idea is based on the ranking on manifold to find the candidate neighborhood and then using estimated local tangent space to select adaptive neighbors. Experimental results demonstrate that the method is not sensitive to the variation of the curvature and sample density on the manifold and can generate satisfying results while other neighborhood selection methods can't get good results.

In this paper, it assumes that data points are sampled directly from the manifold with no noise or little noise. If there have much noise or outliers, the method won't work very well. So to find a robust method to get good embedding result is the emphasis of our future work. Besides, we now select parameter η as an empirical value in our experiments, to find a better method for selecting η is also our future work.

Acknowledgements

This work was supported by the Natural Science Foundation of Guangdong Province of china (07006474, 06300170) and Sci & Tech Research Project of Guangdong Province (2007B010200044), China.

References

[1] J B Tenenbaum, V de Silva, J C Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science, vol.290, no.22, pp.2319-2323, 2000.

[2] S T Roweis, L K Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science, vol.290, no.22, pp.2323-2327, 2000.

[3] M Belkin, P Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. Neural Computation, vol.15, no.6, pp.1373-1396, 2003.

[4] D L Donoho, C E Grimes. Hessian Eigenmaps: Locally Linear Embedding Techniques for High-dimensional Data. Proceedings of the National Academy of Arts and Sciences, 2003.

[5] Z Y Zhang, H Y Zha. Principal Manifolds and Nonlinear Dimensionality Reduction via Tangent Space Alignment. SIAM Journal of Scientific Computing, vol.26, no.1, pp.313-338, 2004.

[6] O Smako, A D Marshall, P L Rosin. Selection of the Optimal Paramater Value for the Isomap Algorithm. Pattern Recognition Letters, vol.27, pp.968-979, 2006.

[7] G H Wen, L J Jiang, N R Shadbolt. Using Graph Algebra to Optimize Neighborhood for Isometric Mapping. International Joint Conference on Artificial Intelligence, 2007.

[8] J Wang , Z Y Zhang, H Y Zha. Adaptive Manifold Learning. Proceedings of 18th Annual Conference on Neural Information Processing System, 2004.

[9] N Mekuz, J K Tsotsos. Paramaterless Isomap with Adaptive Neighborhood Selection. 28th Annual Symposium of the German Association for Pattern Recognition, 2006.

[10] D Y Zhou, J Weston, A Gretton, et al. Ranking on Data Manifolds. Proceedings of 17th Annual Conference on Neural Information Processing System, 2003.

[11] D Y Zhou, O Bousquet, T N Lal, et al. Learning with Local and Global Consistency. Proceedings of 17th Annual Conference on Neural Information Processing System, 2003.

[12] W L Min, K Lu, X F He. Locality Pursuit Embedding. Pattern Recognition, vol.37, pp.781-788, 2004.

[13] M Turk, A Pentland. Face Recognition Using Eigenfaces. IEEE Conf. Computer vision and pattern recognition, 1991.

[14] A M Martinez, A C Kak. PCA Versus LDA. IEEE Trans. Pattern Analysis and Machine Intelligence, vol.23, no.2, pp.228-233, 2001.