Locally Linear Embedding Preserving Local Neighborhood

Tingquan Deng
College of science
Harbin Engineering University
Harbin China

Jinyan Liu, Ning Wang
College of science
Harbin Engineering University
Harbin China

***Abstract*—Dimensionality reduction is an important issue in information processing and has popular applications in many fields, where locally linear embedding (LLE) is widely used due to accuracy and simple to implement. However, LLE is lack of robustness, and sensitive to local structure that can't preserve neighborhood character sometimes. Instead, Laplacian eigenmaps (LE) can overcome these weaknesses. In this paper, a new dimensionality reduction algorithm, called locally linear embedding preserving neighborhood (NLLE) is proposed. It takes the advantage of LLE and LE, and can keep the intrinsic character of high-dimensional data. Several experiments are employed to confirm the effectiveness and robustness of the algorithm.**

Keywords-dimensionality reduction; locally linear embedding; laplacian eigenmaps; preserving local neighborhood.

I. INTRODUCTION

Dimensionality reduction has received more and more attention, which plays an important role in pattern recognition, machine learning, image processing, and data visualization. It refers to mapping high-dimensional manifold into low dimension space, while keeping some features. High dimension manifold can be embedded in a low-dimensional space, and be easy to find useful and important information.

Dimensionality reduction methods generally can be divided into two kinds: global and local [1]. Global approaches include principal component analysis (PCA), and maximum variance unfolding (MVU) [2], etc. However, those methods don't take into account local information and can't unfold this manifold accurately. Local approaches include LLE, LE, and HLLE, etc. They can recover the underlying manifolds successfully. LLE [3] owing to his simple geometrical significance, global optimization, and easy to implement, is widely used. However, LLE [4-7] also has some inherent defects such as easily affected by noise and parameter, and be sensitive to local structure that can't be preserved accurately. LLE can preserve the local linear structure, but can't keep the local neighbor structure in the low dimensional space. Based on above facts, some improvement ideas have been put forward. To overcome noise, weights were introduced to restrain noisy points [8]. The results of LLE are sensitive to the number of neighborhoods and the regularization parameter [9], some methods automated choosing the parameter are proposed [10-11].

Due to the fact that LE is robust to noise and can preserve the neighborhood structure of dataset, the aim of this paper is to combine the advantage of LLE and LE, and present a new method called locally linear embedding preserving neighborhood (NLLE), which can find inherent structure of the manifold and achieve underlying information.

The rest of this paper is organized as follows: Section 2 recalls the classical LLE and LE, Section 3 proposes a new approach to dimensionality reduction called NLLE. Several experiments on artificial datasets and real-world datasets are presented to verify the performance of the proposed algorithm in section 4. Finally concluding comments are included in section 5.

II. LLE, LE AND SLLE

A. LLE

LLE has a simple geometric significance. We assume that the point can be approximated as linear combinations of nearest neighbors. And LLE can preserve the local lineally structure in the low dimensional space. For each point x_i $(i=1,2,\ldots,n)$ who is a vector of size $d\times 1$, choosing k nearby points $N(x_i)$ by Euclidean distance. We assume that x_i can be represented by its nearby points $N(x_i)$ as a linear reconstruction. The linear reconstruction weights can be obtained by minimizing,

$$\sum_{i=1}^{n}\left\| x_i - \sum_{j=1}^{\mathrm{N}(x_i)} w_{ij}x_{ij}\right\|^2 \quad s.t. \quad \sum_{j=1}^{\mathrm{N}(x_i)} w_{ij}=1 \tag{1}$$

978-1-5090-4093-3/16/$31.00 ©2016 IEEE

If x_{ij} belongs to the neighbors of x_i, then we set $w_{ij} \neq 0$, else $w_{ij} = 0$. Mapping x_i into the low dimensional space which preserves the local linear reconstruction weights,

$$\begin{aligned} &\phi(Y) = \sum_{i=1}^{n} \left\| y_i - \sum_{j=1}^{N(y_i)} w_{ij} y_{ij} \right\|^2 \\ &s.t. \ \sum_{i=1}^{n} y_i = 0 \\ &\sum_{i=1}^{n} y_i y_i' = I \end{aligned} \tag{2}$$

The first constrain removes the rotate degree of freedom. While the second constrain requires outputs within a certain range [12]. Using the Lagrange multiplier method, let $M = (I - W)'(I - W)$, (2) turns to $\phi(Y) = tr(Y'MY)$

$$s.t. \begin{cases} \mathbf{1}_{1\times n} Y = \mathbf{0}_{1\times n} \\ \dfrac{1}{n} Y'Y = I_{n\times n} \end{cases} \tag{3}$$

So, it turns to calculate eigenvectors of M corresponding to the smallest $d+1$ eigenvalues [13]. Since the smallest eigenvalue is 0, it's discarded.

B. LE

Laplacian eigenmaps [14] assumes that nearby points in the high dimensional space remain nearby in the low dimensional space. By minimizing

$$\begin{aligned} &\min \ \sum_{i,j=1}^{n} \left\| y_i - y_j \right\|^2 W_{ij} \\ &s.t. \ \sum_{i=1}^{n} y_i = 0 \\ &\sum_{i=1}^{n} \left\| y_i \right\|^2 D_{ii} = I \end{aligned} \tag{4}$$

The weights of nearby points are computed by the heat kernel method. If x_{ij} belongs to the neighbor of x_i, $W_{ij} \neq 0$, else $W_{ij} = 0$. That is,

$$W_{ij} = \begin{cases} \exp(-\dfrac{\left\| x_i - x_j \right\|^2}{\beta}), & if \ x_i \in N_k(x_j) \ \ or \ \ x_j \in N_k(x_i) \\ 0, & otherwise \end{cases} \tag{5}$$

The eigenvectors of $Ly = \lambda Dy$ corresponding to the smallest $d+1$ eigenvalues are the embedding results of high dimensional data in the low dimensional space. The smallest eigenvalue is 0 and is discarded.

C. SLLE

Base on the fact that LLE is often in lack of robustness. C Hou et al propose a stable LLE (SLLE), based on Tikhonov regularization framework，

$$Y^* = \arg\min_{Y \in H_Y} \sum_{i=1}^{N} V(\phi(x_i), y_i) - \lambda D(Y)$$

where V is the loss term, $D(Y)$ is regular term, and λ is a positive real number that adjusts the influence between loss term and regular term. To avoid the influence of noise and the size of neighbors, Hou construct the regular term as:

$$\sum_{i=1}^{N} \sum_{j \sim i} \left\| y_i - y_j \right\|^2$$

Thus, the corresponding objective function is constructed as follows:

$$\arg\min_{Y} \sum_{i=1}^{N} \left\| y_i - \sum_{j \sim i} w_{ij} \right\|^2 - \lambda \sum_{i=1}^{N} \sum_{j \sim i} \left\| y_i - y_j \right\|^2 \tag{6}$$

In order to ensure that the objective function has an optimal solution, a positive real number λ should be chosen to ensure a smaller loss term and a bigger regular term. The optimal problem (6) is solved by the kernel method and semi-definite programs method. In that method, the greatest d eigenvalues are chosen and the corresponding eigenvectors are considered as the embedding low dimensional outputs.

III. LOCALLY LINEAR EMBEDDING PRESERVING LOCAL NEIGHBORHOOD (NLLE)

The SLLE algorithm may be stable to noise and the size of neighbors. But it also has some shortcomings, for instance, the parameter is difficult to be set, if λ is too large, the first part value of (6) is too small, local linear structure isn't preserved, so lead to some local information be lost. If λ is too small, it is little effect of regular term and so lead to the effect of the SLLE is not obvious.

In Fig. 1, point (2, 2) can be lineally reconstructed by points (1, 2), (2, 3), and (1, 3) with weights 1, 1, and -1. However, the neighbors of points (2, 2) are (2.5, 2), (1.5, 2), and (2, 1.5). It presents that local linear structure points and the neighbor points in the low dimensional space are not the same points which are derived by LLE.

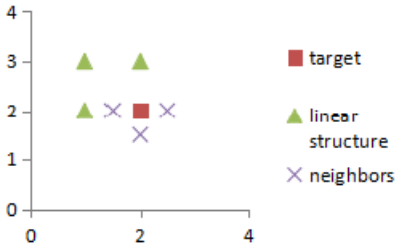

Figure 1. Points present in low-dimension space which are derived by LLE

Fig. 1 shows that the nearby points in the high-dimensional space are mapped to low dimensional space may be not nearby points. Some nearby points may are mapped to the faraway

place. In other words, LLE can remain local linear structures, but the local neighborhood structure can't be preserved. SLLE also doesn't take into account neighborhood structures. So base on above fact, NLLE is proposed, the ideal results can be seen in Fig. 2.

In order to preserving local neighbor structure, we attach a weight w_{ij}' to present the relationship and similarity between neighbor points. However, SLLE doesn't take into account the weight w_{ij}' .

Classical LLE is sensitive to noise. However, LE is robust to noise. So we hope to combine LE with LLE to improve the robustness of LLE.

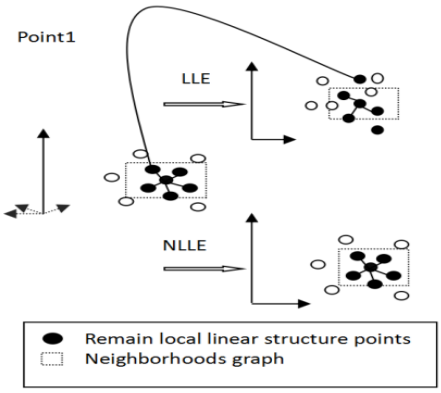

Figure 2. Schematic diagram.

Suppose that a data set contains $x_1, x_2, \ldots, x_n \in R^{D1}$,For each point x_i , we can find its k neighbors through the euclidean distance. Then we need to compute weight w_{ij} which preserves the local linear structure. The reconstruction errors are measured by the cost function,

$$\begin{aligned} &\sum_{i=1}^{n}\left\|x_i-\sum_{j=1}^{\mathrm{N}(x_i)} w_{ij}x_{ij}\right\|^2 \\ &s.t. \quad \sum_{j=1}^{\mathrm{N}(x_i)} w_{ij}=1 \end{aligned} \tag{7}$$

subject to a constrain $w_{ij}=0$ if x_{ij} is not k-neighbor of x_i .

For each point x_i ,

$$\begin{aligned} \left\|x_i-\sum_{j=1}^{\mathrm{N}(x_i)} w_{ij}x_{ij}\right\|^2 &= \sum_{j,l=1}^{\mathrm{N}(x_i)} w_{ij}'\left(x_i-x_{ij}\right)'\left(x_i-x_{il}\right)w_{il} \\ &= w_i' G w_i \end{aligned} \tag{8}$$

where G is Gram matrix with $k\times k$ elements, the k neighbors weight value of x_i is denoted by w_i

$$G_{jl}=\left(x_i-x_{ij}\right)'\left(x_i-x_{il}\right) \tag{9}$$

we can obtain,

$$w_i=\frac{G^{-1}\mathbf{1}}{\mathbf{1}'G^{-1}\mathbf{1}} \tag{10}$$

The weight w_{ij}' to represent the neighborhood relationship and similarity between point x_i sand x_j which is defined by the heat kernel method as follows,

$$w_{ij}'=\begin{cases}\exp(-\dfrac{\left\|x_i-x_j\right\|^2}{\beta}), & if\ x_i\in N_k(x_j)\ \ or\ \ \ x_j\in N_k(x_i)\\ 0, & otherwise\end{cases} \tag{11}$$

If x_{ij} belongs to the neighbor of x_i , then we set $w_{ij}'\neq 0$, else $w_{ij}'=0$.

Assume the mapping function f .

$$f: R^{D1}\rightarrow R^{d} \quad d<D1$$

maps x_i to y_i in the low dimensional space, whose k neighbors are presented by $N(y_i)$.

The embedding coordinate y_i who is a vector of size $1\times d$, in the low space can be obtained by minimizing the cost function,

$$\begin{aligned} \arg\min_{Y}\ &(1-\rho)\times\sum_{i=1}^{n}\left\|y_i-\sum_{i=1}^{\mathrm{N}(y_i)} w_{ij}y_{ij}\right\|^2+\rho\times\sum_{i=1}^{n}\sum_{j=1}^{\mathrm{N}(y_i)} w_{ij}'\left\|y_i-y_{ij}\right\|^2 \\ s.t. \quad &\sum_{i=1}^{N} y_i=0 \\ &\sum_{i=1}^{N} D_{ii}'\left\|y_i\right\|^2=I \end{aligned} \tag{12}$$

where ρ is a balance parameter which is limited from zero to one. The effectiveness of ρ is a compromise between LLE and LE. The first constrain ensures the vector rotation invariance in the low dimensional space and the second one ensures the uniqueness of the output variable. We can assume that,

$$\varepsilon=(1-\rho)\times\sum_{i=1}^{n}\left\|y_i-\sum_{i=1}^{\mathrm{N}(y_i)} w_{ij}y_{ij}\right\|^2+\rho\times\sum_{i=1}^{n}\sum_{j=1}^{\mathrm{N}(y_i)} w_{ij}'\left\|y_i-y_{ij}\right\|^2$$

and let $L'=D'-W'$, $M_{N\times N}=(I-W)'(I-W)$, $D_{ii}'=\sum_{j=1}^{k} w_{ij}'$,

where D' is a diagonal, $W=\left[w_{ij}\right]_{n\times n}$, $W'=\left[w_{ij}'\right]_{n\times n}$.

We can assume that $H=(1-\rho)\times M+\rho\times L'$, H is a matrix of size $n\times n$, and obtain that,

$$\varepsilon = (1-\rho)\times tr(Y'MY) + \rho \times tr(Y'L'Y) = tr(Y'HY) \tag{13}$$

In view of constraining condition, employing the Lagrange multiplier method, we obtain.

$$\varepsilon = tr(Y'HY) + tr(1 - Y'DY) \tag{14}$$

In order to minimize ε , it turns into calculate the generalized eigenvectors of $HY = \lambda D'Y$ corresponding to the smallest $d+1$ eigenvalues. Since the smallest eigenvalue is 0 or close to 0 , any vector can be eigenvectors corresponding to 0 , so we discard the first eigenvalue to ensure the uniqueness of the output variable.

IV. EXPERIMENTS

A. Experiments on Artificial Datasets

We produce an artificial dataset to verify our method. It is generated by the following function,

$$t = (3\pi/2)\times(1+2\times rand(1,N));$$
$$X = [t\cos(t); 21\times rand(1,N); t\sin(t)];$$

Which is called Swiss roll dataset, Let $N = 800$, we implement both LLE and NLLE on Swiss roll whose dimension is 3. In the low dimensional embedding of the dataset, which is derived from LLE in Fig. 3 (a), (b), we choose one point (magenta) as the target. The red points preserve the local linear points, but they are not nearby points when mapped to the low dimensional space (green points are the neighbor points in the low dimensional space). The ideal effect is that the red points and green points are the same. However, In Fig. 3 (b), (d), the embedding result by using NLLE shows that the red points and green points are the same. In other words, NLLE can preserve the local linear structure as well as the local neighbor structure.

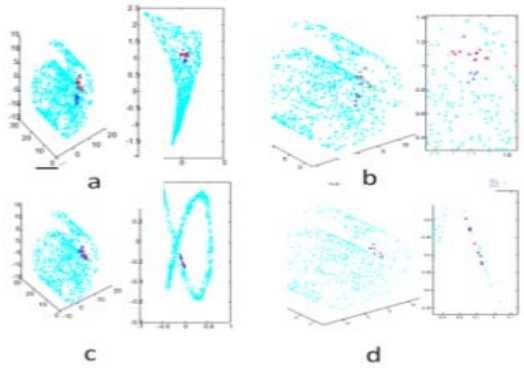

Figure 3. (a)Embedding of LLE，(b)A local spot of (a). (c)Embedding of NLLE. (d)A local spot of(c)

We choose another artificial dataset, called helical line to verify the robustness of our proposed algorithm. We add white Gaussian noise to the helical line data. We use LLE, WKLLE, SLLE, and NLLE to test the helical line. We unfold this manifold, the ideal curve is rising that presents the length from the first point to our target point. In Fig. 4, we can observe that LLE, WKLLE and SLLE can't unfold the manifold accurately. Because some points have the same output coordinates. It means that some points overlap with each other in the low dimension space. The embedding derived by NLLE is more close to ideal results. So NLLE is more robust to noise.

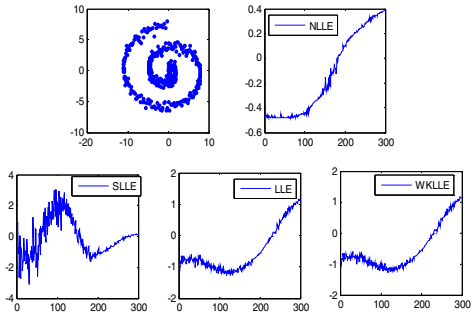

Figure 4. Embedding derived by the NLLE, SLLE, LLE, and WKLLE on helical line with noise.

B. Experiments on Real-world Datasets

We employ the Mnist handwritten digits set and the Breast cancer sets to evaluate the effectiveness of our proposed algorithm.

1) Mnist handwritten digits set

The set contains 60000 training examples and 10000 test examples of digits '0'-'9'. Each sample is an image with the resolution 28*28. We extract 200 samples from handwritten digits '0', '1', '2', '3', and '4' training sets respectively, totally 1000 samples. The 10-fold cross-validation method is applied to evaluate the effectiveness of our algorithm.

We can split data into 10 sets and every set contain 10% of total samples, one sets as the test set, and others as training sets. We repeat 10 times and take the average accuracy as the accuracy of the algorithm.

The new NLLE is compared against the classical LLE, WKLLE, and SLLE. The experimental results are shown in Fig.5 and TABLE I. The dimension of the output space is set to be 2 , The neighbor size k is changed from 8 to 50. $q = 0.9$, $\lambda = 0.0116$ in SLLE.

TABLE I. shows that the classification accuracy of the proposed algorithm (NLLE) is a little lower than other algorithms when k is small. But the classification accuracy of the proposed algorithm (NLLE) is higher than other algorithms significantly when k is changing, and NLLE more stable to k . In most datasets, NLLE has high classification accuracy than other algorithms significantly.

LLE is sensitive to neighbor size. If k is too large It can't unfold the manifold and some local information is lost. SLLE is very sensitive to λ , but our method is not so strict to the parameter, because our method is keeping neighbor structure no matter what the parameter is. SLLE pays attention to the stability of the algorithm to the parameters k and the noise,

but our algorithm more attention to the classification accuracy in the low dimensional space. So experiments verify that NLLE has higher classification accuracy than SLLE.

TABLE I COMPARATIVE RESULTS OF DIFFERENT ALGORITHM WITH DIFFERENT SIZES OF NEIGHBORS ON THE MNIST DATASET(1000 POINT)(%)

k	LLE	WKLLE	SLLE	NLLE
8	84.27	**84.68**	31.63	83.91
10	**84.60**	84.51	33.64	84.05
20	74.54	78.81	36.59	**86.58**
30	63.96	66.41	24.13	**85.37**
40	64.18	62.21	25.85	**84.71**
50	36.79	63.23	23.12	**84.22**

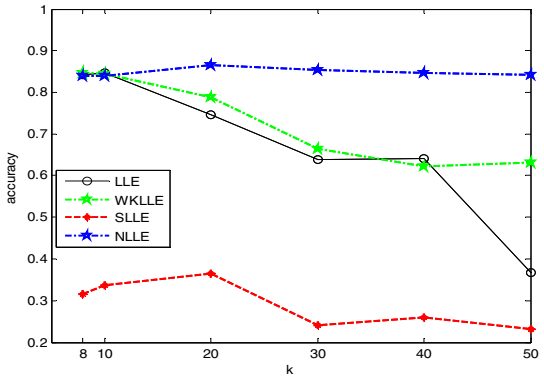

Figure 5. Comparative results of different algorithm with different sizes of neighbors on the Mnist dataset (1000 points)

We extract 800 samples from handwritten digits '0', '1', '2', '3', '4' training sets, respectively, totally 4000 samples to test the classification accuracy of the proposed algorithm. It is seen from TABLE II that NLLE has high classification accuracy, no matter how much the amount of data is. Furthermore, NLLE is more stable when the sizes of neighbors are changed. SLLE spends a lot of times, when we experiment on Minst dataset SLLE use more than 5 hours, however we just need less than 3 minutes, $\lambda = 0.0008$ in SLLE.

TABLE II COMPARATIVE RESULTS OF DIFFERENT ALGORITHM WITH DIFFERENT SIZES OF NEIGHBORS ON THE MNIST DATASET (4000 POINT)(%)

k	LLE	WKLLE	SLLE	NLLE
8	91.20	90.75	63.55	**97.06**
10	91.13	90.62	62.12	**95.49**
20	83.68	84.38	74.00	**90.52**
30	42.29	72.23	75.52	**91.86**

2) Breast cancer wisconsin data set

This dataset contains 699 samples. Because some values are lost, those samples are deleted. We employ 683 samples, containing 9 attributes with each sample. It is classified into two classes, malignant and benign. TABLE III presents the classification accuracies by using the 10-fold cross-validation. The 2- dimension embedding are derived by LLE, WKLLE, SLLE and NLLE. The parameter k is ranged from 30 to 60, and $q = 0.9$, $\lambda = 0.1759 \times 10^{-4}$ in SLLE. It is found that the classification accuracy of the proposed algorithm (NLLE) is higher than other algorithms significantly, and more stable.

TABLE III COMPARATIVE RESULTS OF DIFFERENT ALGORITHM WITH DIFFERENT SIZES OF NEIGHBORS ON CANCER DATASET (683 POINTS) (%)

k	LLE	WKLLE	SLLE	NLLE
30	95.48	93.37	54.17	**96.42**
40	75.36	77.32	44.73	**96.51**
50	73.00	73.00	54.49	**96.32**
60	89.65	89.75	47.23	**96.17**

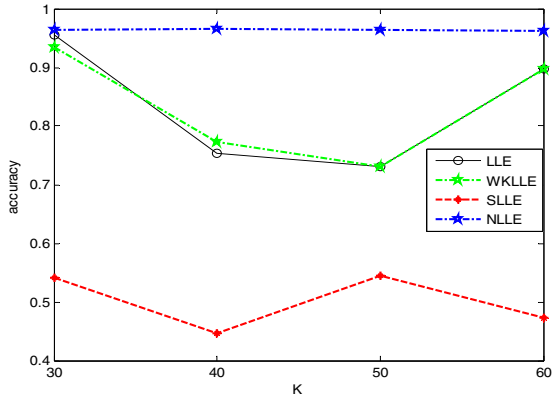

Figure 6. Comparative results of different algorithm with different sizes of neighbors on breast cancer dataset (683 points)

We delete the duplicates and employ 449 instances, which contain 9 attributes of each sample. $q = 0.5$, $d = 2$. $\lambda = 0.0499 \times 10^{-4}$ in SLLE. From the TABLE IV, the classification accuracy of the proposed algorithm (NLLE) is higher than other algorithms when k is changed. It is also shows that NLLE has high classification accuracy, no matter how much the amount of data is.

TABLE IV COMPARATIVE RESULTS OF DIFFERENT ALGORITHM WITH DIFFERENT SIZES OF NEIGHBORS ON BREAST CANCER DATASET (449 POINTS)(%)

k	LLE	WKLLE	SLLE	NLLE
10	87.33	87.06	59.40	**94.90**
15	85.59	89.20	50.60	**94.92**
20	86.53	84.61	48.44	**94.68**
25	91.56	91.65	47.83	**94.43**
30	93.03	93.07	45.06	**94.43**

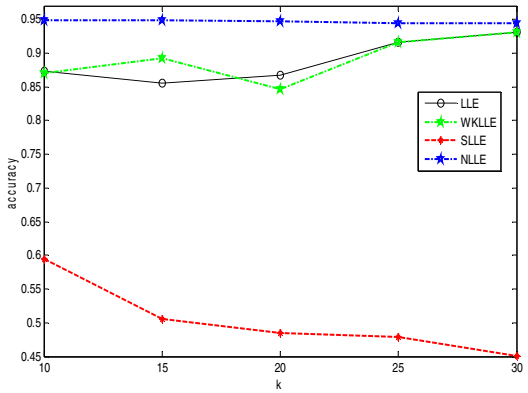

Figure 7. Comparative results of different algorithm with different sizes of neighbors on breast cancer dataset (449 points)

C. Parameters Discussions

From the above experimental results, we need to discuss the selection of parameters

1) The choice of k (neighbors size)

LLE is sensitive to noise, when k is too large, LLE can't reflect local property, and the results more tend to the effectiveness of PCA [15]. When k is small, LLE separates or breaks into sub manifolds by mistake, which leading to not preserve the topological structure of dataset in the low dimensional space. The choice of k is very strict, which reflects the linear structure. It is known that the choice of neighbor size for LE is not strict unlike LLE. Only when k is very small, it may cause neighborhood not connected, if k is too large, it may cause a short circuit，for uniform data, k can be changed within a certain range, the dimension reduction results by LE are fairly stable than LLE. So NLLE is more stable than LLE. .

In the following experiments, we adapt 3 datasets. Iris dataset contains 150 instances and each instance has 4 attributes. Breast cancer dataset contains 569 instances and each instance has 30 attributes. Sheffield database contains 400 instances and each instance has 92*112 attributes. k is ranged from 4 to 30. The results show that NLLE is more stale when k is changed within a certain range for most of the data sets.

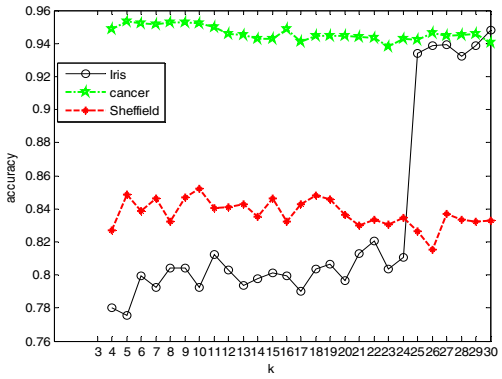

Figure 8. Comparative results of different dataset with different sizes of neighbors by using NLLE

2) The choice of d (d-dimensional embedding)

There are some methods to estimate the intrinsic dimension of datasets [16]. It has a big effect on unfolding the manifold accurately. Nevertheless, the intrinsic dimension of dataset is not easy to be exactly estimated and generally given by empirism or practical demand.

In the following experiments, we adapt 3 datasets. Breast cancer dataset contains 569 instances and each instance has 30 attributes. Mnist handwritten digit set contains 1000 instances and each instance has 784 attributes. Sheffield database contains 400 instances and each instance has 92*112 attributes. d is ranged from 2 to 30 and k is set to be 8.

The experimental result shows that classification accuracy is affected by the changing of the parameter d, different dataset is needed to find the intrinsic dimension to unfold the manifold accurately.

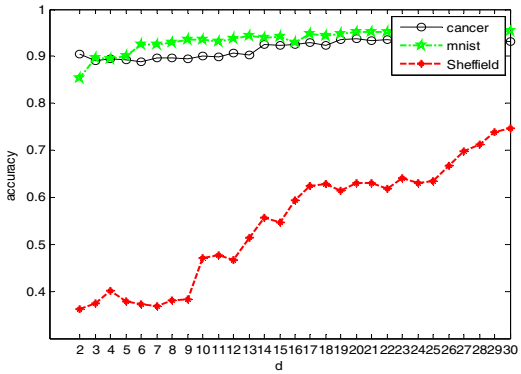

Figure 9. Comparative results of different dataset with different sizes of dimension by using NLLE

3) The choice of ρ

When $\rho = 0$, NLLE has the same effectiveness as LLE, and when $\rho = 1$, NLLE has the same effectiveness as LE. In practice, a compromise value of ρ is selected according to empirism or practical demand. Some datasets is not sensitive to ρ. Because it preserve the local neighbor relationship no matter what ρ is. However, SLLE is very sensitive to a positive real number λ. If λ is a negative number, may be SLLE is similar to NLLE. However, it is not in accordance with Tikhonov regularization framework and can't find the optimal solution. Even though the objective functions are similar between NLLE and SLLE, they have different starting points and focuses. The focus of SLLE is stable to noise and the size of neighbors, but NLLE is high classification accuracy.

V. CONCLUSION

This paper proposes a new dimensional reduction method called local neighborhood preserving LLE. The new algorithm can preserve locally linear structure, and maximally remain local neighbor structure in the low dimensional space. NLLE can unfold the manifold accurately and robust to noise. The experimental results confirm the validity of the proposed algorithm.

Future work is to apply the proposed algorithm to outlier detection in the high-dimension space. And it is also needed to work on how to estimate intrinsic dimension and the compromise parameter to unfold the manifold accurately.

ACKNOWLEDGMENT

This paper is supported by the National Natural Science Foundation of China with Grant 11471001. And this paper is funded by the International Exchange Program of Harbin Engineering University for Innovation-oriented Talents Cultivation.

REFERENCES

[1] C. Hou, C. Zhang, Y. Wu, "Stable local dimensionality reduction approaches," Pattern Recognition, 2009, 42(9), pp.2054-2066.

[2] J. Sun, P. Diaconis , "The Fastest Mixing Markov Process on a Graph and a Connection to a Maximum Variance Unfolding Problem," Siam Review, 2004, 48(4), pp.681-699.

[3] S. T. Roweis, L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science, 2000, 290(5500), pp.2323-2326.

[4] H. Chang, D. Y. Yeung, "Robust locally linear embedding," Pattern Recognition, 2006, 39(6), pp.1053-1065.

[5] J. Yin, D. Hu, Z. Zhou , "Noisy manifold learning using neighborhood smoothing embedding," Pattern Recognition Letters, 2008, 29(11), pp.1613-1620.

[6] G. Daza-Santacoloma, C. D. Acosta-Medina, G. Castellanos-Domínguez , "Regularization parameter choice in locally linear embedding," Neurocomputing, 2010, 73(10–12), pp.1595-1605.

[7] L. Feng, Z. Weijie, G. Suicheng, "Local Linear Laplacian Eigenmaps: A Direct Extension of LLE," Pattern Recognition Letters, 2016, 75, pp.30-35.

[8] B. Niu, "Weighted Kernel Locally Linear Embedding for Dimensionality Reduction," Journal of Information & Computational Science, 2014, 11(7), pp.2109-2116.

[9] J. Chen, Y. Liu, "Locally linear embedding: a survey," Artificial Intelligence Review, 2011, 36(1), pp.29-48

[10] A. Álvarez-Meza, J. Valencia-Aguirre, G. Daza-Santacoloma, "Global and local choice of the number of nearest neighbors in locally linear embedding," Pattern Recognition Letters, 2011, 32(16), pp.2171-2177.

[11] J. Wang , "Real local-linearity preserving embedding," Neurocomputing, 2014, 136(8), pp.7-13.

[12] L. K. Saul, S. T. Roweis, "Think globally, fit locally: unsupervised learning of low dimensional manifolds," Journal of Machine Learning Research, 2003, 4(2), pp.119-155.

[13] R. Hettiarachchi, J. F. Peters, "Multi-manifold LLE learning in pattern recognition," Pattern Recognition, 2015, 48(9), pp.2947-2960.

[14] M. Belkin, P. Niyogi, "Laplacian Eigenmaps for dimensionality reduction and data representation," Neural Computation, 2003, 15(6), pp.1373-1396.

[15] J. B. Tenenbaum, S. V. De, J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," Science, 2000, 290(5500), pp.2319-2323.

[16] M. Y. Fan, Q. Hong, Z. Bo, "Intrinsic dimension estimation of manifolds by incising balls," Pattern Recognition, 2009, 42(5), pp.780–787