

# Research on Image Dimension Reduction Algorithm Based Manifold Learning

Yuanshao Hou<sup>1, a</sup> and Yao Zhang<sup>1, b</sup>

<sup>1</sup>Zhengzhou University of Industrial Technology, Zhengzhou 451100, China

<sup>a</sup>houyuanshao123@126.com, <sup>b</sup>304963405@qq.com

**Keywords:** Manifold learning; Feature extraction; Intrinsic dimensionality; Data mining

**Abstract.** For 1D and 2D image feature extraction ignore the structural information of the image, resulting the loss of recognition accuracy, the feature extraction of 3D and multiplanar images while considering the data structure with each other, but the curse of dimensionality increases the computational complexity. Using manifold learning, embedding stable manifold into the original data space, so that the multidimensional data in the feature data is mapped to the manifold, discovered that the low dimensional structure hidden in high dimensional data which people unable to perceive, and then under the premise of without losing the data information, reduce the dimension of the raw data, so as to reduce the computational complexity.

## Manifold and Manifold Learning

**The Basic Principle of Manifold Learning.** The six elements of manifold learning are: high dimensional data observation, select the appropriate intrinsic dimension and the low dimensional representation of high dimension data, the embedded mapping of the manifold, refactoring mapping in data recovery, and the limitation conditions of the filter noise. Manifold learning assumption there is a low dimensional manifold in the high dimensional data, the low dimensional manifold can contain most of the information of the data. Through the mapping and perception, the original high dimensional data is mapped to the low dimensional embedding space, and the low dimensional embedding space retains the original data correlation and geometry property [1]. Assuming that high-dimensional observation data  $x = \{x_i \in R^D, i=1, L, N\}$ , of which N is the data capacity for a given sample, D is the decision of the high - dimensional data. Assumes that the low dimensional manifold is  $\Omega^d \subseteq R^d$ , intrinsic dimension is  $d$  ( $d < D$  and  $d = D$ ). The image dimension reduction of manifold learning is to find the mapping relation from high dimensional data to the low dimensional embedding space.  $f$ , so as to  $y_i = f(x_i)$ , thus obtain the low dimensional representation  $y_i \in R^d$  of the high dimensional data  $x_i$  [2]; In data reconstruction recovery phase, in order to restore the original image to the maximum precision and preserve the geometric structure of the original manifold data, according to the mapping function  $y_i = f(x_i)$  to meet the specific constraints, thereby restore the original data  $x_i = f^{-1}(y_i)$ .

**The Classification of Manifold Learning.** Manifold learning can have different classification methods based on different constraints and feature extraction criteria. The most commonly used is to classify them according to the structural characteristics of manifolds:

(1) Linear analysis method: Linear refers to the mapping function  $f$  to meet the linear relationship, including the main analysis method, multi scale transform, linear discrimination, etc. Such as two dimensional principal component analysis (2DPCA) method, the main idea is that in a row or column of the original data  $x$  to find a underdetermined linear mapping matrix  $y$ ,  $y = \varphi_1 x \varphi_2$  which line mapping  $\varphi_1$  and  $\varphi_2$  are non random underdetermined matrix. Therefore, compression ranks of the rows and columns of a matrix  $y$  is smaller than that of the  $x$  [3].

(2) Nonlinear analysis method: Linear analysis is a special form of nonlinear analysis, its main disadvantage is that when the data dimension is very high, the calculation of the characteristic

vector is not feasible. Due to the uncertainty of data distribution and dispersion, nonlinear analysis method on the assumption that the data in low dimensional nonlinear manifold, and then from the local neighborhood structure of the data, transform the manifold learning problem into the optimization problem of the objective function, by solving the objective function, it is concluded that the manifold data need to maintain the geometric properties, which is the underdetermined matrix, and then obtain the low dimensional embedding representation of the original high-dimensional data[4]. Typical local linear embedding method (LLE) is shown in Fig. 1 Firstly, for each data point  $i=1,2,L,N$  of the sample data  $x_i$ , Define  $k$  nearest neighbor  $x_{ij}$  in its neighborhood, the reconstruction error is defined as:

$$\varepsilon(w) = \sum_i \left\| x_i - \sum_j x_{ij} w_j^i \right\|_2 \quad (1)$$

Among them,  $W_{ij}$  weights between  $x_i$  and  $x_{ij}$ , when  $\sum_{j=1}^k w_j^i = 1$ , the local optimal reconstruction matrix can be calculated by lagrange operator

$$w_j^i = \frac{\sum_{m=1}^k (Q^i)^{-1}_{jm}}{\sum_{p=1}^k \sum_{q=1}^k (Q^i)^{-1}_{pq}} \quad (2)$$

$Q^i$  is a singular matrix, the conditions for mapping high dimensional data to low dimensional space are:

$$\min \varepsilon(Y) = \sum_{i=1}^N \left\| y_i - \sum_{j=1}^k w_j^i y_{ij} \right\|^2 \quad (3)$$

$y_i$  are the output vector of  $x_i$ ,  $\varepsilon(Y)$  is the value of loss function for data reconstruction. The nonlinear analysis method preserves the geometric structure of the neighborhood, and the rotation, translation, and scaling of the neighbor points can also be well mapped. Therefore, the nonlinear analysis method can maintain the invariance of manifold structure [5].

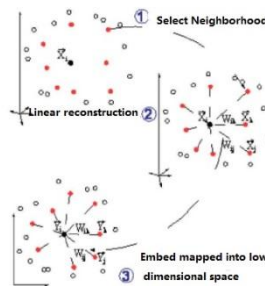


Figure 1. Local linear embedding representation

### Intrinsic Dimension Estimation

The intrinsic dimension estimation is divided into two: one is using a geometric method to estimate and the local geometry of the high-dimensional data is examined; another is the eigenvalue method that is suitable for linear data, which is on global intrinsic dimension estimation [6]. In view of the characteristics of manifold learning, the advantage of the geometric method to estimate the intrinsic dimension is more and more concerned by people [7].

**K - near Neighbor Graph Method.** Assume that there are the independent random samples

$y_n = \{Y_1, \dots, Y_n\}$  in space  $R^D$ , with  $y_n$  as a vertex, for any point to connect all its  $K$  - nearest neighbour, get a  $K$  - neighbor graph. Grouping all  $Y_i$   $K$ - nearest neighbour as  $N_{k,i} = N_{k,i}(y_n)$ , can be got the total length of  $K$ - neighbor graph for internal edge:

$$L_{r,k}(y_n) = \sum_{i=1}^n \sum_{Y \in N_{k,i}} |Y - Y_i| \quad (r > 0 \text{ for eighting constant}) \quad (4)$$

Assump that compact  $m$  dimensional manifold  $(M, g)$ , and through the evaluate limit of the total length of the internal line

$$\lim \frac{L_{r,k}(y_n)}{n(d' - r)/d'} = \begin{cases} \infty & d' < m \\ \beta_{m,r,k} \int_M f^\alpha(y) \mu_g(dy) & d' = m \\ 0 & d' > m \end{cases} \quad (5)$$

Among them  $\beta_{m,r,k}$  is constant, which has nothing to do with  $f$ , when and only then  $d' = m$ , meet  $\alpha = (m - r)/m$

Let  $l_n = \log L_{r,k}(y_n)$ , By approximation conditions

$$l_n = a \log n + b + \varepsilon_n \quad (6)$$

There  $a = (m - r)/m$ ,  $b = \log \beta_{m,r,k} + r/m H_\alpha^{(M,g)}(f)$ ,  $\varepsilon_n$  is variance.

$H_\alpha^{(M,g)}$  is the limit value of limited value, and nonzero:

$$H_\alpha^{(M,g)}(f) = \frac{1}{1 - \alpha} \log \int_M f^\alpha(y) \mu_g(dy) \quad (7)$$

By sorting out the above formulas and concepts, we get the estimation of the intrinsic dimension of the data set is:

$$\hat{m} = \text{round} \left\{ r / \left( 1 - \hat{a} \right) \right\} \quad (8)$$

There,  $\hat{a}$  is the linear mean square estimation of  $a$ ,  $r > 0$  is weighting constant.

**Maximum Likelihood Estimation Method.** Sampling random sample  $X_1, \dots, X_n$  in the space  $R^D$ , set a point  $x$  as the center of them, circle with  $R$  as radius, then the number of samples  $X_1, \dots, X_n$  in the ball are:

$$N(t, x) = \sum_{i=1}^n I \{ X_i \in S_x(t) \} \quad (0 < t < R) \quad (9)$$

The distance between the  $k$ th nearest neighbor  $x$  to the  $x$  in the sampled sample  $X_1, \dots, X_n$ :

$$\frac{k}{n} \approx f(x) V(m) [T_k(x)]^m \quad (10)$$

For non-stationary process of fixed  $t$ , the parameter of Poisson process is:

$$\lambda(t) = f(x) V(m) m t^{m-1} \quad (11)$$

The likelihood function is established for the Poisson process:

$$L(m, \theta) = \int_0^R \log \lambda(t) dN(t) - \int_0^R \lambda(t) dt \quad (12)$$

Partial derivatives of likelihood function is calculated, and the maximum likelihood estimation are obtained:

$$\frac{\partial L}{\partial \theta} = \int_0^R dN(t) - \int_0^R \lambda(t) dt = 0 \quad (13)$$

$$\frac{\partial L}{\partial m} = 0 \quad (14)$$

Intrinsic dimension can be got from the above formulas:

$$\hat{m}_R(x) = \left[ \frac{1}{N(R, x)} \sum_{j=1}^{N(R, x)} \log \frac{R}{T_j(x)} \right]^{-1} \quad (15)$$

### Manifold Learning Algorithm Based on Sparse Representation

**Sparse Representation.** Sparse representation theory pointed out that if the image data set of information in a manifold or in some kind of transformation is sparse (i.e. vector zero elements in the majority), then by the sparsity of the image data were underdetermined random mapping, image compression data without having to worry about losing any information [8]. Set image  $\mathcal{X}$  as  $M \times N$  matrix, converts the matrix into column vector  $x^{(vec)} \in R^{M \times N}$ , then the vector feature  $y^{(vec)} \in R^{M^{(vec)}}$  of the sample form of one dimensional compressed image can be obtained by the following equation:

$$y^{(vec)} = \Phi x^{(vec)} \quad (16)$$

There,  $\Phi \in R^{M^{(vec)} \times (MN)}$  ( $M^{(vec)} < MN$ ) is an underdetermined random matrix satisfying the RIP condition (random distribution of Gaussian distribution). Sparse representation of data acquisition mode is shown in Fig. 2.

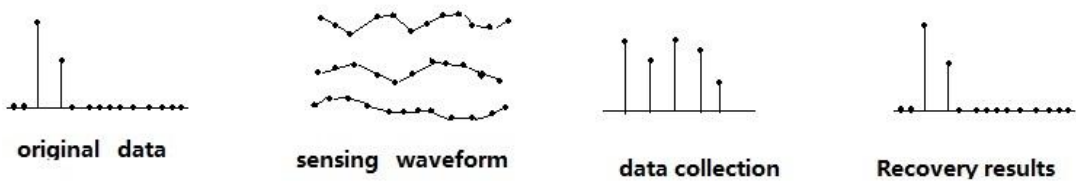


Figure 2. Sparse representation of data collection

$$f \times [\varphi_1, \varphi_2, \dots, \varphi_m]^T = y \xrightarrow{l_1 \min} \hat{f} \quad (17)$$

**Sparse Local Embedding Method.** To solve these two problems, the local linear embedding method based on sparse representation (SLLE) is proposed [9]. SLLE by modifying the LLE to the local area of the modeling objective function, the introduction sparsity constraint, after sparse constraint optimization of local linear embedding, can be used in the high dimensional space to optimize the obtained (sparse) weight in the low dimensional space reconstruction of samples, and then the manifold structure of the sample is reduced [10].

Sparse local embedding algorithm:

1. Input  $N \times m$  dimensional experimental samples  $A = [Y_1, Y_2, \dots, Y_N] \in R^{m \times N}$ , error  $\varepsilon > 0$ .
2.  $l_2$  norm optimization is carried out for the experimental samples  $A = [Y_1, Y_2, \dots, Y_N] \in R^{m \times N}$  to

generate a random matrix  $\Phi \in R^{m \times m}$ ,  $Y = \Phi A$ , generating a new set of samples  $A = [Y_1, Y_2, \dots, Y_N] \in R^{m \times N}$ .

3. Set  $\hat{A}_i = [Y_1, K Y_{i-1}, Y_N]$ , transformation to 1 norm optimization problem:

$$\hat{W}_i = \arg \min \left\| \hat{W}_i \right\|_1 \quad s.t. \quad \hat{A}_i \hat{W}_i = Y_i, \sum_j W_{ij} = 1$$

4. For the sample set repeat step 3, the weight vector of the first sample  $i$  is:  $W_i = [W_{i1}, W_{i2}, \dots, W_{in}]$ , the weight of the first  $i$  sample is  $W_{ij} = 0$ .

5. Fixed the weight solution and minimizing the error of low dimensional neighborhood

$$\text{reconstruction } \varepsilon(x) = \sum_i \left| X_i - \sum_j W_{ij} X_j \right|_2$$

6. Final output:  $X = \arg \min \varepsilon(X)$

## Acknowledgements

Project of the National Natural Science Foundation of China (NO. 6115200), a supervised image classification research based on two dimensional random mapping and one norm optimization.

## References

- [1] Davenport M.A Theoretical Analysis of Joint Manifolds (Technical Report TREE - 0901) [J]. Rice University ECE Department, 2009, 116(24): 118-121.
- [2] Yang J. Two - Dimensional PCA: A New Approach to Appearance - Based Face Representation and Recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 26(1): 131 - 137.
- [3] Hegde C M, Wakin R, Baraniuk. Random Projections for Manifold Learning [M]. New York: Neural Information Processing Systems NIPS, 2007: 641 - 648.
- [4] Baraniuk RG, Wakin MB. Random Projections of Smooth Manifolds [R]. New York: Foundations of Computational Mathematics, 2009: 35.
- [5] Lian-wei Zhao, Si-wei Luo, Yan-chang Zhao. Research on low dimensional embedding and embedding dimension of high dimensional data [J]. Journal of software, 2005, 16(8): 1423-1430.
- [6] Xu Song - hua, Lau F, Cheung W K, et al. Automatic generation of artistic Chinese calligraphy [J]. IEEE Intelligent Systems, 2005, 20(3): 32-39.
- [7] Gh.A. Shirali. A new method for quantitative assessment of resilience engineering by PCA and NT approach: A case study in a process industry [J]. Reliability Engineering & System Safety, 2013, 119:88 - 94.
- [8] Roweis S T, Saul L K. Nonlinear Dimensionality Reduction by Locally Linear Embedding [J]. Science, 2010, 290(55): 23-26.
- [9] Zhang Zhenyue, Zhang Hongyuan. Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment [J]. Sci. Computer. SIAM, 2004, 26(1): 319-338.
- [10] Li Yang. Building k-edge-connected Neighborhood Graphs for Distance-based Data Projection [J]. Pattern Recognition Letters, 2005, 26(13): 2015-2021.