A Neighborhood Adaptive Semi-supervised Local Fisher Discriminant Analysis Algorithm Postprint

Authors: Du Wei, Fang Liqing, Qi Ziyuan

Date: 2018-05-20T00:00:00+00:00

Abstract

To address the limitation that globally uniform setting of local neighborhood size based on empirical experience fails to capture the differences in local geometric structures when applying localization ideas to discriminant analysis of multimodal data, we propose a Neighborhood Adaptive Semi-Supervised Local Fisher Discriminant Analysis (NA-SELF) algorithm. Building upon the semi-supervised local Fisher discriminant analysis framework, the algorithm combines Mahalanobis distance and cosine similarity to determine the initial number of nearest neighbors, which is further adjusted according to sample space probability density estimation. The feature dimensionality reduction performance of the proposed algorithm is evaluated on artificial datasets and five UCI benchmark datasets, and compared with typical dimensionality reduction algorithms and discriminant analysis algorithms employing traditional #1; nearest neighbor methods. Experimental results demonstrate the superior effectiveness of the proposed algorithm.

Full Text

A Neighborhood Adaptive Semi-Supervised Local Fisher Discriminant Analysis Algorithm

Du Wei, Fang Liqing, Qi Ziyuan
(Dept. of Artillery Engineering, Ordnance Engineering College, Shijiazhuang 050003, China)

Abstract: When applying localization ideas to discriminant analysis of multimodal data, the global uniform setting of local neighborhood size based on empirical experience fails to capture differences in local geometric structure. To address this limitation, this paper proposes a Neighborhood Adaptive Semi-Supervised Local Fisher Discriminant Analysis (NA-SELF) algorithm. Building upon the semi-supervised local Fisher discriminant analysis framework, the algorithm determines the initial neighbor count by combining Mahalanobis distance and cosine similarity, then adjusts the neighbor count adaptively according to probability density estimation in the sample space. The feature dimensionality reduction performance of the proposed algorithm is validated using synthetic datasets and five UCI benchmark datasets, with comparisons against typical dimensionality reduction algorithms and discriminant analysis algorithms employing traditional k-nearest neighbor methods. Experimental results demonstrate the superior effectiveness of the proposed algorithm.

Key Words: local neighborhood; adaptive; semi-supervised local Fisher discriminant analysis; dimensionality reduction

0 Introduction

With the development of information technology, data encountered in many research and application domains often suffer from high dimensionality, redundancy, and information mixing. To avoid the "curse of dimensionality," improve efficiency, and fully exploit the essential information in raw data, effective dimensionality reduction is necessary. As an important data preprocessing technique, dimensionality reduction has been widely applied in image processing, pattern recognition, and computer vision.

Dimensionality reduction algorithms map data from high-dimensional space to low-dimensional space through linear or nonlinear transformations while preserving structural characteristics. These algorithms are categorized as supervised or unsupervised based on whether sample labels are utilized. Linear Discriminant Analysis (LDA) represents a typical supervised method, while Principal Component Analysis (PCA) exemplifies unsupervised approaches. However, unsupervised methods ignore label guidance, whereas supervised methods require numerous labeled samples, which is often cost-prohibitive in practice. Semi-supervised dimensionality reduction methods, which leverage both unlabeled and limited labeled data, have achieved excellent results.

To address LDA's poor performance on multimodal data, many researchers have introduced localization ideas to mine data manifold structure using local information, such as Locality Preserving Projection (LPP), Local Fisher Discriminant Analysis (LFDA), and Marginal Fisher Analysis (MFA). Existing algorithms typically construct neighborhoods through global uniform parameter settings based on experience, ignoring differences in local geometric structure and thereby affecting the separability of low-dimensional projection vectors. Consequently, automatically determining neighborhood size according to distance metrics between data points has become a worthwhile research problem.

Based on this analysis, this paper proposes a Neighborhood Adaptive Semi-Supervised Local Fisher Discriminant Analysis (NA-SELF) algorithm. The algorithm combines distance and angular similarity metrics to construct neighborhoods and employs sample space probability density estimation to adaptively adjust neighbor counts, effectively overcoming the limitation of global uniform neighborhood parameters in the SELF algorithm. Finally, the low-dimensional vectors obtained through NA-SELF are fed into a Support Vector Machine (SVM) for recognition, validating the algorithm' s effectiveness.

1 Semi-Supervised Local Fisher Discriminant Analysis

Sugiyama et al. proposed the Semi-Supervised Local Fisher Discriminant Analysis (SELF) algorithm by effectively fusing LFDA and PCA. LFDA improves multimodal data processing by describing local sample information but tends to overfit when labeled samples are insufficient. PCA, conversely, can capture global distribution using unlabeled samples. The SELF algorithm combines these advantages, possessing both LFDA's ability to utilize category information for dimensionality reduction and PCA' s capacity to obtain global distribution without category information.

Assuming a sample set containing D-dimensional features and c categories, denoted as $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^{N}$, where labeled samples are $\{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$ with category labels $y_i \in \{1, 2, \dots, c\}$. The global scatter matrix is defined as:

$$\mathbf{S}^{(t)} = \frac{1}{2}\sum_{i,j=1}^{N} W_{ij}^{(t)}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top$$

where the weight $W_{ij}^{(t)} = 1/N$. The local between-class scatter matrix $\mathbf{S}^{(lb)}$ and local within-class scatter matrix $\mathbf{S}^{(lw)}$ of LFDA can be defined in the following pairwise forms:

$$\mathbf{S}^{(lb)} = \frac{1}{2}\sum_{i,j=1}^{N} W_{ij}^{(lb)}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top$$

$$\mathbf{S}^{(lw)} = \frac{1}{2}\sum_{i,j=1}^{N} W_{ij}^{(lw)}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top$$

where the weight matrices are:

$$W_{ij}^{(lb)} = \begin{cases} A_{ij}(1/n_l - 1/n_{l'}) & \text{if } y_i = l, y_j = l' \neq l \\ 1/n_l & \text{if } y_i = y_j = l \\ 0 & \text{otherwise} \end{cases}$$

$$W_{ij}^{(lw)} = \begin{cases} A_{ij}/n_l & \text{if } y_i = y_j = l \\ 0 & \text{otherwise} \end{cases}$$

Here, n_l is the number of samples in class l, and the similarity matrix $\mathbf{A}$ describes similarity between samples $\mathbf{x}_i$ and $\mathbf{x}_j$, with $A_{ij} \in [0,1]$. The similarity matrix can be defined in various forms, such as Gaussian similarity, k-nearest neighbor similarity, and local scaling similarity.

From equations (1)-(3), the between-class and within-class scatter matrices for SELF are defined as:

$$\mathbf{S}^{(b)} = \beta \mathbf{S}^{(lb)} + (1-\beta)\mathbf{S}^{(t)}$$

$$\mathbf{S}^{(w)} = \beta \mathbf{S}^{(lw)} + (1-\beta)\mathbf{I}$$

where the weight coefficient $\beta \in [0,1]$ is a standard matrix. This weight coefficient enables the algorithm to possess characteristics of both LFDA and PCA, with flexibility increased by adjusting its value. Clearly, SELF is equivalent to PCA when $\beta = 0$ and equivalent to LFDA when $\beta = 1$. Finding the optimal projection transformation matrix $\mathbf{T}$ involves solving the following maximization objective function:

$$\mathbf{T} = \underset{\mathbf{T} \in \mathbb{R}^{D \times d}}{\operatorname{argmax}} \operatorname{tr}\left((\mathbf{T}^\top \mathbf{S}^{(w)} \mathbf{T})^{-1} \mathbf{T}^\top \mathbf{S}^{(b)} \mathbf{T}\right)$$

The solution of the transformation matrix in equation (9) is equivalent to solving the generalized eigenvector problem in equation (10):

$$\mathbf{S}^{(b)} \alpha = \lambda \mathbf{S}^{(w)} \alpha$$

Thus, the transformation matrix $\mathbf{T}$ consists of the generalized eigenvectors $\{\alpha_i\}_{i=1}^d$ corresponding to the d largest generalized eigenvalues from equation (10).

2 Neighborhood Adaptive Semi-Supervised Local Fisher Discriminant Analysis

Traditional neighbor count setting methods generally fall into two categories: k-nearest neighbor and ϵ-nearest neighbor. The SELF algorithm employs the ϵ-nearest neighbor method when calculating the local scaling similarity matrix $\mathbf{A}$, with globally uniform parameters set empirically. However, actual collected sample data often exhibits differences in local geometric structure, meaning different samples require different neighbor sample sets during low-dimensional mapping, which impacts algorithm performance differently. To address this issue, this paper adopts a neighborhood parameter adaptive adjustment method to improve algorithm robustness while enhancing the discriminability of low-dimensional features.

2.1 Similarity Measurement

The SELF algorithm describes local scaling by calculating the Euclidean distance between a sample point and its k-th nearest neighbor. However, Euclidean distance only measures spatial position between samples and cannot reflect the overall set structure. Mahalanobis distance is unaffected by feature scale selection, while cosine similarity measures similarity using the cosine of the vector angle. Therefore, to fully reflect sample similarity, this paper combines cosine similarity and Mahalanobis distance:

$$d_{ij}^c = \cos(\mathbf{x}_i, \mathbf{x}_j) \cdot d_{ij}^m$$

where d_{ij}^m and d_{ij}^c are the Mahalanobis distance and cosine similarity between samples, respectively, and:

$$d_{ij}^m = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{M}^{-1} (\mathbf{x}_i - \mathbf{x}_j)}$$

$$d_{ij}^c = \frac{\mathbf{x}_i^\top \mathbf{x}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|}$$

Equation (11) fuses spatial position and angle information between sample points, equivalent to adding an influence factor with range $[0, 1]$ to Mahalanobis distance. The smaller the angle between two vectors, the smaller the influence factor, and the smaller d_{ij}^c.

Based on the advantage of combining Mahalanobis distance and cosine similarity to reflect data distribution, the similarity matrix elements in the SELF algorithm are described as:

$$A_{ij} = \exp\left(-\frac{d_{ij}^c}{2\sigma_i \sigma_j}\right)$$

where σ_i is the local scale of sample point $\mathbf{x}_i$, defined as the distance to its k-th nearest neighbor. The initial neighbor count k is determined by the mean of all sample similarity coefficients. If the similarity coefficient A_{ij} is greater than $\bar{A}$, then $\mathbf{x}_j$ is a neighbor of $\mathbf{x}_i$. Clearly, the neighbor count for each sample obtained through this method may differ.

2.2 Neighborhood Parameter Adaptive Adjustment

When constructing neighborhoods, samples with similar features tend to be densely distributed, while samples with poor similarity are more sparsely distributed. If the neighbor count k can be adjusted adaptively according to the probability density of local region sample points, the low-dimensional features obtained through dimensionality reduction can better reflect the essential structure of original data. Parzen window probability density estimation is a non-parametric method that does not require assumptions about the probability density function form but estimates the overall probability density from the data itself. Therefore, Parzen window estimation is applied to neighborhood construction for adaptive adjustment of the initial neighbor count k determined by similarity mean $\bar{A}$.

Assuming $\mathcal{X}$ is a D-dimensional space containing dataset $\mathcal{D}$, for $\mathbf{x} \in \mathbb{R}^D$, the Parzen window probability density estimate is:

$$p(\mathbf{x}) = \frac{1}{Nh^D} \sum_{i=1}^{N} \phi\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right)$$

where $V = h^D$ is the window volume, N is the number of dataset samples, h is the window width, and $(\mathbf{x} - \mathbf{x}_i)$ is the distance between $\mathbf{x}$ and $\mathbf{x}_i$ calculated according to equation (11). $\phi(\mathbf{x})$ is the window function satisfying $\phi(\mathbf{x}) \geq 0$ and $\int \phi(\mathbf{x}) d\mathbf{x} = 1$. The Gaussian window function with good smoothness is selected as the window function. The window width h is set to the distance between $\mathbf{x}_i$ and its k-th nearest neighbor $\mathbf{x}_{ik}$.

Let k_i be the initial neighbor count for data point $\mathbf{x}_i$. The neighborhood probability density of data point $\mathbf{x}_i$ is:

$$p_i = \frac{1}{Nh^D} \sum_{\mathbf{x}j \in \mathcal{N}_i} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2k_i^2}\right)$$

where $\mathcal{N}_i$ represents the neighbor set of $\mathbf{x}_i$. The average neighborhood probability density of all samples in the dataset is:

$$\bar{p} = \frac{1}{N} \sum_{i=1}^{N} p_i$$

The neighbor parameter is then adjusted according to:

$$k_i' = \text{floor}\left(k_i \cdot \frac{p_i}{\bar{p}}\right)$$

where floor is the floor function.

Analysis of equation (15) shows that when the probability density near data point $\mathbf{x}_i$ exceeds the average, the neighbor count k_i automatically increases, making distant data pairs contribute less to dimensionality reduction; conversely, k_i automatically decreases, making nearby data pairs contribute more, thereby preserving local neighborhood structure and facilitating recovery of global structural information in the low-dimensional dataset.

2.3 Neighborhood Adaptive Semi-Supervised Local Fisher Discriminant Analysis Algorithm Flow

The Neighborhood Adaptive Semi-Supervised Local Fisher Discriminant Analysis (NA-SELF) algorithm proceeds as follows:

Input: D-dimensional space data sample set $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^{N}$, where the number of labeled samples is l, and the target dimension of low-dimensional feature space is d ($d \leq D$).

Output: Projection transformation matrix $\mathbf{T}$ and low-dimensional feature vectors $\mathbf{Y} = \mathbf{T}^\top \mathbf{X}$.

a) Calculate similarity coefficients A_{ij} between data points in high-dimensional space using equation (12), and obtain the initial neighbor count k_i for each sample from the mean of similarity coefficients $\bar{A}$.

b) Calculate the neighborhood probability density p_i of samples using Parzen window probability density estimation, adjust the neighborhood parameter k_i according to equation (15), construct the similarity matrix $\mathbf{A}$, substitute into equations (4)-(5) to obtain weight matrices $\mathbf{W}^{(lb)}$ and $\mathbf{W}^{(lw)}$, and consequently derive the local between-class scatter matrix $\mathbf{S}^{(b)}$ and local within-class scatter matrix $\mathbf{S}^{(w)}$.

c) Solve for the generalized eigenvectors corresponding to the d largest generalized eigenvalues according to equation (10) to obtain the projection transformation matrix $\mathbf{T}$, thereby acquiring the low-dimensional space data representation.

2.4 Algorithm Time Complexity Analysis

The time complexity difference between the proposed NA-SELF algorithm and the original algorithm primarily manifests in steps a) and b) of the NA-SELF algorithm flow, namely calculating the similarity matrix and adaptively adjusting neighborhood parameters. Assuming the total number of data samples is N and the original feature dimension is D, the time complexity for calculating cosine similarity and Mahalanobis distance using equation (11) is $O(DN^2)$. Determining the initial neighbor count has complexity $O(N^2)$. The time complexity for calculating neighborhood probability density and adjusting neighborhood parameters using equations (14)-(15) is also $O(N^2)$.

Let the time complexity of the original SELF algorithm throughout its entire flow be O_{SELF}. After simplification, the time complexity of NA-SELF becomes:

$$O_{\text{NA-SELF}} = O_{\text{SELF}} + O(DN^2) + O(N^2)$$

According to equation (16), the complexity difference between the improved and original algorithms mainly relates to the total sample count and original feature dimension—larger sample sizes and feature dimensions yield greater time complexity. Therefore, in practical pattern recognition applications, data attributes should be carefully considered. The proposed method demonstrates excellent applicability for multimodal data with relatively small sample sizes and feature dimensions. Additionally, in scenarios where recognition accuracy is prioritized over computational efficiency, the proposed algorithm can be effectively employed for multimodal data dimensionality reduction.

3 Experiments and Analysis

To verify the dimensionality reduction performance of the proposed algorithm, experiments were implemented using MATLAB R2013a on an Intel Core i5-3470 CPU @ 3.20 GHz with 4 GB RAM.

3.1 Artificial Data Experiments

In these experiments, PCA, LFDA, SELF, and NA-SELF algorithms were applied to binary artificial datasets for dimensionality reduction, with visualization experiments providing intuitive performance comparison. In each artificial data experiment, 200 data points were randomly generated from bivariate normal distributions, with each class containing 100 unlabeled and 10 labeled samples. The two classes are represented by circles and triangles, while unlabeled and labeled samples are represented by hollow and solid markers, respectively. Figures 1 [Figure 1: see original paper] through 3 [Figure 3: see original paper] show one dataset from Experiments 1-3 and the projection directions obtained by different algorithms, where lines represent one-dimensional projection spaces drawn with different line styles. In each experiment, one dataset served as training samples and another as test samples. The test samples were first dimensionality-reduced to obtain the projection transformation matrix, which was then applied to the test samples. Low-dimensional features were input to an SVM for training and recognition, with 100 trials conducted. The mean recognition rates are shown in Table 1 . The weight coefficient β in SELF and NA-SELF algorithms was set to 0.5, the neighbor count k in SELF was set to 7, the SVM kernel function was radial basis function with penalty parameter $C = 1$ and kernel parameter $g = 1$.

Experiment 1: The binary dataset shown in Figure 1 has one modality per class. Unlabeled sample means are at $(4, 0)$ with covariance matrix $[4, 0; 0, 4]$, while labeled sample means are at $(4, 3)$ with covariance matrix as a second-order identity matrix. The correct projection direction is clearly horizontal. Results show that PCA and NA-SELF obtained good projection directions. LFDA, influenced by the lower labeled samples, shows significant deviation. Since SELF utilizes both labeled and unlabeled samples, its projection direction lies between those of PCA and LFDA.

Experiment 2: The binary dataset shown in Figure 2 [Figure 2: see original paper] has one class with one modality and another with two modalities. The central class has mean $(0, 0)$, while the two side classes have means at $(4, 0)$ and $(-4, 0)$. Unlabeled sample covariance matrix is $[1, 0; 0, 10]$, with labeled sample means and covariance matrices identical to unlabeled samples. The correct projection direction is horizontal. NA-SELF obtained a good projection direction, while PCA chose the vertical direction that maximizes dataset variance, and LFDA selected a direction with larger squared distance between different classes, resulting in vertical projection. Consequently, PCA and LFDA show significant deviation.

Experiment 3: The binary dataset shown in Figure 3 [Figure 3: see original paper] has two modalities per class. Unlabeled sample means are at $(8, 4)$ and $(-8, -4)$ with covariance matrix $[2, 0; 0, 2]$, while labeled sample means are identical to unlabeled samples with covariance matrix as a second-order identity matrix. The correct projection direction is vertical. Since the distance between same-class modalities exceeds that between different classes in the same modality, PCA selects a horizontal projection direction. LFDA' s projection direction shows some deviation because distant same-class samples contribute less to projection direction selection. NA-SELF, through adaptive neighbor count adjustment, obtains a similarity matrix that more fully reflects the local structure of sample data, thus achieving better projection direction.

Table 1 shows that in Experiment 1 with unimodal data, NA-SELF' s average recognition accuracy is slightly lower than PCA' s. However, in Experiments 2 and 3 with multimodal data, NA-SELF achieves higher average recognition accuracy than other algorithms, with the highest average across all three experiments, demonstrating clear advantages and better applicability for multimodal data dimensionality reduction.

3.2 UCI Data Experiments

Five standard datasets from the UCI Machine Learning Repository were selected for dimensionality reduction, with low-dimensional features input to an SVM for classification. The UCI datasets used are shown in Table 2 .

Table 2. UCI Dataset Information

Dataset	Classes	Features	Training Samples	Test Samples
Ionosphere	2	34	200	151
Wine	3	13	120	58
Vehicle	4	18	600	246
Segment	7	19	1500	810
Satimage	6	36	3000	1435

For comparison, PCA, LFDA, SELF, and NA-SELF algorithms were evaluated. In SELF, neighbor count $k = 7$, and parameter β was obtained through 5-fold cross-validation from $\{0.1, 0.3, 0.5, 0.7, 0.9\}$. SVM parameters remained as in Section 3.1. In training samples, the ratio of labeled to unlabeled samples was randomly assigned as 1:3. Taking the Wine dataset reduction results as an example, Figure 4 [Figure 4: see original paper] shows the 3D spatial distribution of the first three vectors of low-dimensional feature sets from training samples when reduced to 5 dimensions.

Analysis of Figure 4 reveals that PCA shows poor dimensionality reduction performance with severe mixing between different class feature sets. LFDA, utilizing only limited labeled samples, also exhibits some mixing between classes. SELF simultaneously uses numerous unlabeled and limited labeled samples, achieving basic class separation after dimensionality reduction. NA-SELF employs the combined Mahalanobis distance and cosine similarity method, which reflects both spatial position and angle information between sample points, yielding more precise similarity and consequently better dimensionality reduction.

Figure 5 [Figure 5: see original paper] shows the recognition accuracy of test samples on the Wine dataset across different dimensionality reduction dimensions. To compare the impact of different similarity measures, NA-SELF based on Euclidean distance and NA-SELF based on combined Mahalanobis distance and cosine similarity are also compared. The results demonstrate that different dimensionality reduction algorithms and dimensions produce varying recognition accuracies, with NA-SELF achieving the highest classification precision within a certain range.

Each algorithm was tested 100 times on all five datasets. The average recognition accuracies are shown in Table 3 , where values represent the mean of highest accuracies per experiment (standard deviation in parentheses). Direct classification using original unreduced data (None) is also included for comparison.

Table 3. Average Recognition Accuracy of Each Algorithm (%)

Dataset	None	PCA	LFDA	SELF	NA-SELF (Euclidean)	NA-SELF
Ionosphere	72.11(2.36)	93.39(3.38)	90.33(1.44)	66.95(2.03)	68.82(0.89)	74.13(2.52)
Wine	78.32(2.02)	69.72(1.28)	92.78(2.14)	90.02(1.58)	61.26(1.79)	90.83(1.50)
Vehicle	68.88(2.55)	69.13(3.14)	77.16(2.53)	73.92(2.03)	92.93(2.52)	92.96(2.12)
Segment	66.73(2.26)	78.76(2.87)	81.25(2.25)	64.45(3.01)	78.70(2.96)	80.75(2.85)
Satimage	78.32(2.02)	69.72(1.28)	92.78(2.14)	90.02(1.58)	81.25(2.25)	82.63(2.29)

The results indicate that original data without dimensionality reduction contains considerable redundant information, resulting in lower recognition rates than most dimensionality-reduced versions. PCA demonstrates good stability but, as a linear method, ignores nonlinear data structure. LFDA may overfit with insufficient labeled samples. Consequently, PCA and LFDA recognition rates are mostly lower than SELF. Since SELF employs globally uniform neighborhood parameters, its recognition accuracy and stability are lower than NA-SELF

(Euclidean). The proposed algorithm' s similarity measure more fully reflects inter-sample similarity, achieving optimal accuracy on three of five datasets and the highest average across all datasets.

To compare time complexity between NA-SELF and original SELF, Table 4 lists average test times and the percentage increase of the improved algorithm over the original.

Table 4. Test Time Comparison (ms)

Dataset	SELF	NA-SELF	Increase
Ionosphere	45.2	59.7	32.11%
Wine	38.6	51.4	33.07%
Vehicle	128.5	161.2	25.48%
Segment	245.8	342.5	39.32%
Satimage	512.3	786.1	53.43%

The test time is slightly longer, with the percentage increase growing with test sample quantity and feature dimension, indicating that data attributes significantly impact algorithm improvement. Therefore, in practical pattern recognition applications, data characteristics should be carefully considered. The proposed method shows excellent applicability for multimodal data with relatively small sample sizes and feature dimensions. Moreover, in scenarios prioritizing recognition accuracy over computational efficiency, the algorithm can be effectively applied to multimodal data dimensionality reduction.

4 Conclusion

This paper proposes a Neighborhood Adaptive Semi-Supervised Local Fisher Discriminant Analysis algorithm. The algorithm combines Mahalanobis distance and cosine similarity to describe inter-sample similarity and employs Parzen window probability density estimation for adaptive neighbor count adjustment during neighborhood construction, effectively avoiding arbitrary manual selection and providing better expression of local geometric structure characteristics. Experiments on synthetic datasets and five UCI benchmark datasets demonstrate that compared with typical PCA, LFDA, and SELF algorithms, NA-SELF obtains superior projection spaces and more discriminative low-dimensional feature vectors. The similarity measure combining Mahalanobis distance and cosine similarity proves more effective than Euclidean distance-based methods. However, in semi-supervised dimensionality reduction algorithms, weight coefficient values are currently obtained through cross-validation. Developing methods for rapid acquisition of effective weight coefficients represents a future research direction.

References

[1] Xie Jun, Liu Jian. A new local discriminant projection method [J]. Chinese Journal of Computers, 2011, 34(11): 2243-2250.

[2] Zhai L D, Ding Z Y, Jia Y, et al. A word position-related LDA model [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2011, 25(6): 909-925.

[3] Nguyen V H, Golinval J C. Fault detection based on Kernel Principal Component Analysis [J]. Engineering Structures, 2010, 32(11): 3683-3691.

[4] Yang W Y, Liang W, Xin L, et al. Subspace Semi-supervised Fisher Discriminant Analysis [J]. Zidonghua Xuebao//Acta Automatica Sinica, 2009, 35(12): 1513-1519.

[5] Su Zuqiang, Tang Baoping, Liu Ziran, et al. Fault diagnosis based on orthogonal semi-supervised local Fisher discriminant analysis [J]. Journal of Mechanical Engineering, 2014, 50(18): 7-13.

[6] Yang Xiyang, Deng Chaoyang, Li Zhiwei. Semi-supervised fuzzy Fisher dimensionality reduction analysis [J]. Journal of Xiamen University: Natural Science, 2015, 54(6): 869-875.

[7] Yu J B. Bearing performance degradation assessment using locality preserving projections [J]. Expert Systems with Applications, 2011, 38(6): 7440-7450.

[8] Sugiyama M. Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis [J]. Journal of Machine Learning Research, 2007, 8: 1027-1061.

[9] Yan S, Xu D, Zhang B, et al. Graph Embedding And Extension: A General Framework For Dimensionality Reduction [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2006, 29(1): 40-51.

[10] Zhang Xiaotao, Tang Liwei, Wang Ping, et al. Adaptive neighborhood construction manifold learning algorithm and fault dimensionality reduction diagnosis [J]. Journal of Vibration, Measurement & Diagnosis, 2016, 36(6): 1210-1215.

[11] Sugiyama M. Semi-supervised local Fisher discriminant analysis for dimensionality reduction [J]. Journal of Machine Learning Research, 2010, 78(1-2): 35-61.

[12] Liu Zhiyong, Yuan Yuan. Semi-supervised enhancement based on geodesic distance [J]. Computer Engineering and Applications, 2011, 47(21): 202-204.

[13] Li Wenbo, Wang Dayi, Liu Chengrui. A quantitative evaluation method for fault diagnosability of a class of nonlinear systems [J]. Journal of Astronautics, 2015, 36(4): 455-462.

[14] Yang Wangcan, Zhang Peilin, Wu Dinghai, et al. Fault diagnosis based on improved semi-supervised locality preserving projection algorithm [J]. Journal

of Central South University: Science and Technology, 2015, 46(6): 2059-2064.

[15] Liu Han, Zhang Qing, Meng Lihua, et al. Comprehensive alarm method for equipment condition based on Parzen window estimation [J]. Journal of Vibration and Shock, 2013, 32(3): 110-114.

Note: Figure translations are in progress. See original paper for figures.