PAPER • OPEN ACCESS

Comparative study of different dimensionality reduction methods in hyperspectral image classification

To cite this article: Lei Kang *et al* 2021 *J. Phys.: Conf. Ser.* **2024** 012009

View the article online for updates and enhancements.

Comparative study of different dimensionality reduction methods in hyperspectral image classification

Lei Kang[1a], Xiaoqing Hu[1*], Chengcheng Zhong[1], Kai Zhang[1] and Yanan Jiang[2]

[1]School of Science, China University of Geosciences (Beijing), Beijing, 100083, China

[2]School of Mathematical Sciences, Beijing Normal University, Beijing, 100875, China

[a]2119190050@cugb.edu.cn [*]Corresponding author's e-mail: xiaoqinghu_321@163.com

Abstract. Considering the high-dimensional characteristics of hyperspectral image (HSI) data, researchers generally adopt the dimensionality reduction (DR) methods to reduce the complexity and computing time of subsequent classification or regression tasks while preserving the intrinsic structure information of the data. At present, the research on DR of HSIs data mostly focuses on the application performance of a single method in specific tasks and few studies have been conducted on the adaptability of different DR methods to HSIs data. From the perspective of spectral domain and spatial domain, this paper makes a comparative study on the performance of various linear and nonlinear DR methods in the task of HSI classification of aerial HSI of Matiwan Village in Xiongan New Area. Specifically, it includes principal component analysis, independent component analysis, isometric mapping, Laplacian eigenmaps, autoencoder, etc. The results show that the intrinsic structure in Xiongan HSI is mainly linear structure. Compared with the nonlinear DR methods, the linear DR methods can better preserve the intrinsic structure information of data at a lower time-consuming cost. For the subsequent classification tasks, the linear DR methods have better classification performance and are more suitable for HSI data.

1. Introduction

With the development of remote sensing technology, the ability of imaging spectrometers to capture spectral and spatial information is becoming more and more powerful. Meanwhile, with the expression of features is becoming more refined, the dimension of spectral features increases simultaneously. The high dimensionality with hundreds of bands imposes several challenges to hyperspectral images (HSIs) classification, such as Hughes phenomenon and high demands of computational resources[1].

In previous studies, researchers generally adopted the dimensionality reduction (DR) methods to reduce the complexity of data and preserve the desirable intrinsic structure information of data without sacrificing significant information in order to deal with the problem caused by the high-dimensional characteristics of HSI data[2]. According to whether the supervision information was needed, DR methods could be broadly divided into supervised methods and unsupervised ones. The former, including linear discriminant analysis (LDA)[3], etc, could reduce the dimensionality of HSIs under the situation with limited labelled training samples. In order to make full use of the intrinsic structure information contained in the whole data, many unsupervised DR methods, including principal component analysis (PCA)[4], independent component analysis (ICA)[5], isometric mapping (Isomap)[6], locally linear embedding (LLE)[7], Laplacian eigenmaps (LE)[8] and autoencoder

(AE)[9], were widely used in HSIs DR tasks. In addition, some semisupervised DR methods were proposed for HSIs DR tasks and they achieved better DR based on both unsupervised and supervised information, such as semisupervised dimensional reduction based on sparsity preserving[2].

Although many DR methods have been proved to be effective in HSIs DR tasks, cumulative variance, reconstruction error or classification accuracy cannot fundamentally explain the effectiveness of the DR methods. The core of HSIs DR is to compress the dimension of features while preserving the intrinsic linear and nonlinear structure information of data, that is, the degree of retention of intrinsic structure information of data is the key to measure the DR effect and whether the DR method is suitable for HSIs data. However, there is still a lack of research on the adaptability of DR methods to structural information in HSI data. In this work, we took the aerial HSI of Matiwan Village in Xiongan New Area[10] (hereinafter referred to as Xiongan data set) as an example and reduced the dimension by using linear and nonlinear DR methods. We discussed the characteristics of HSIs data and its adaptability to different DR methods by comparing the DR time, classification accuracy of the data after DR and the retention of data structure information. Finally, we provided a basis for the selection of DR methods for HSIs.

2. Methods

DR refers to a method of using the low-dimensional embedding feature $\boldsymbol{Z} \in \mathbb{R}^d$ to represent the original feature $\boldsymbol{X} \in \mathbb{R}^D$ through feature mapping on the premise of preserving the important information and intrinsic structure of the original data X, where $d < D$. According to different feature mapping methods, DR methods can be categorized into linear DR methods and nonlinear DR methods.

2.1. Linear dimensionality reduction methods

The linear DR methods maps the original features to low-dimensional space by linear projection and takes the low-dimensional representation as the new features. PCA[4] is the most basic unsupervised linear DR method. According to the maximum variance theory or minimum error theory, it projects original features to the direction where the data changes most. KPCA[11] introduces the kernel projection methods to PCA. The former employs the kernel functions to calculate the covariance matrix, maps original features to high-dimensional space and then uses PCA to map the data to low-dimensional space. Different kernel functions have different low dimensional representations. KPCA employing linear kernel function (KPCA-linear) still belongs to linear DR method, while KPCA employing nonlinear kernel function (KPCA-rbf) belongs to nonlinear DR method. SVD[12] is a linear DR method which is similar to PCA. Different from the latter, there is no need to centralize the data when using SVD to reduce the dimension of data. ICA[5] achieves the purpose of DR by decomposing the original feature into several statistically independent components to maximize the independence of each component.

2.2. Nonlinear dimensionality reduction methods

Among the nonlinear DR methods, the widely used ones are kernel-based nonlinear DR methods, manifold learning DR methods and autoencoder. KPCA is a typical kernel-based nonlinear DR method. Isomap[6], LLE[7] and LE[8] are typical manifold-based nonlinear DR methods. Isomap is isometric mapping algorithm which introduces the concept of neighborhood graph and holds that samples are only connected with their adjacent samples. For distant data points, the distance can be calculated through the shortest path. On this basis, DR and distance preservation were carried out. The core idea of LLE is that each data point can be approximately reconstructed by linear combination of multiple adjacent points. And then the high-dimensional data is projected into the low-dimensional space to preserve the local linear reconstruction relationship among data points. LE is similar to LLE algorithm. It constructs the relationship among data points from a local perspective and hopes data points related to each other will be as close as possible in the low-dimension space. T-SNE[13] is also a nonlinear DR method, which is suitable for visualization of high-dimensional data after DR and can

reconstruct its data trend in low dimensions based on original trend of data. Autoencoder[9] is a feedforward neural network which is trained to approximate an identity mapping function. The network consists of Encoder and Decoder. Encoder maps features from high-dimensional space to low-dimensional space and Decoder maps features from low-dimensional space to high-dimensional space. The network is optimized by minimizing reconstruction error between the input of Encoder and the output of Decoder.

3. Experimental Results and Analysis

The experiment was conducted over Xiongan data set[10]. Considering that the spectral bands of HSI data contain rich spectrum information and a number of studies showed that considering spatial neighborhood information could improve the performance of HSI classification. So, we discussed the effects of different DR methods in spectral domain and spatial domain respectively and studied the adaptability of HSI data set to different DR methods.

Table 1. Classification accuracy of different DR methods in spectral domain (%).

Classifier	Linear DR Methods				Nonlinear DR Methods				
	PCA	KPCA -linear	ICA	SVD	KPCA -rbf	Isomap	LLE	LE	AE
LR	84.0	83.9	44.6	**84.3**	74.1	61.6	48.9	55.6	48.0
MLP	**89.2**	88.8	51.3	88.4	79.1	66.1	53.9	56.7	54.0
RF	87.8	87.8	53.2	**88.3**	79.9	69.5	62.8	67.1	55.8

3.1. Experimental Data Set

The spectral range of Xiongan HSI is 400—1000nm, with 256 bands and a spatial resolution of 0.5 m. The image size is 3750×1580 pixels. The land cover types labelled here are 20, which are mainly cash crops. Figure 1 present the available ground truth information of Xiongan HSI. In order to reduce the amount of experimental calculation, 2000 samples were selected randomly to form the samples set *U* and each category contains 100 samples. The positions of samples selected were represented by black dots in Figure 1.
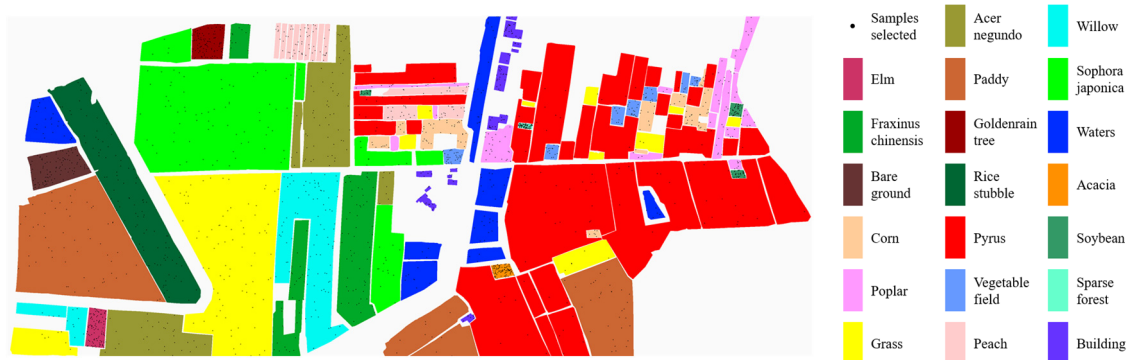

Figure 1. The ground-truth map of Xiongan HSI.

3.2. Experimental Settings

For the data after spectral domain standardization, nine DR methods were considered to reduce the dimension of Xiongan data set in spectral domain and spatial domain respectively, four of which were linear DR methods, such as PCA, KPCA-linear, ICA, SVD. In addition, five nonlinear DR methods were applied, such as KPCA-rbf, Isomap, LLE, LE and AE. The number of spectral bands was reduced to 15 which was empirically chosen as a trade-off between computational complexity and accuracy with minimum information loss.
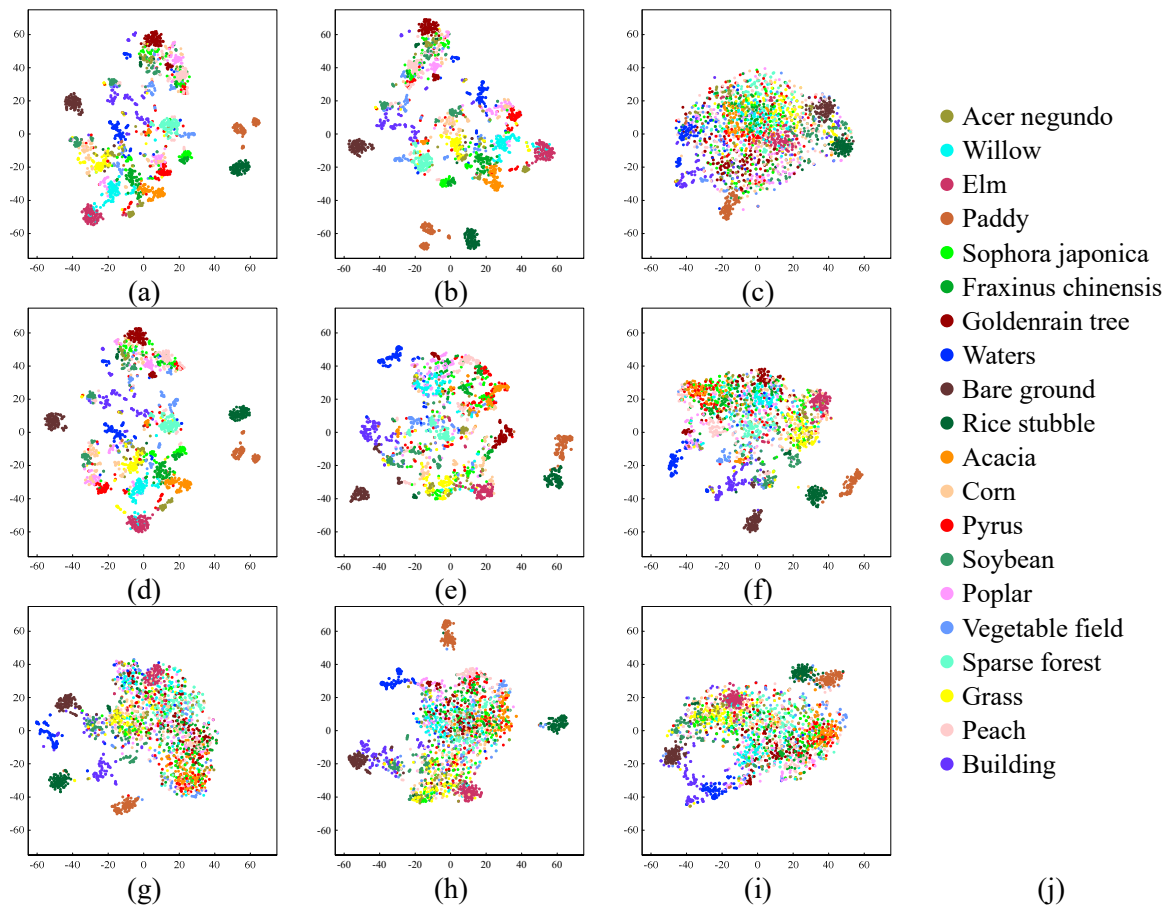
Figure 2. Visualization of data points distribution after DR in spectral domain. (a) PCA. (b) KPCA-linear. (c) ICA. (d) SVD. (e) KPCA-rbf. (f) Isomap. (g) LLE. (h) LE. (i) AE. (j) Legend.

Specifically, in the process of DR in spectral domain, the selected data points might have position offset and could not represent the typical characteristics of the current class of data points. In order to address this problem, we first took the data points to be classified as the center for neighborhood sampling on the data after spectral domain DR. Then, we combined the mean and variance of the features in each band of the current neighborhood data points as the features. In the spatial domain DR process, we sampled the neighborhood of the data points after the spectral domain DR, and then reduced the dimension after the neighborhood feature flatten operation. The neighborhood size was set to 21 uniformly.

In addition, the number of Encoder nodes in AE network was set to d -512-256-128, where d was the dimension number of input data and the structure of Decoder to be symmetrical with Encoder. AE network was constructed by full connection between all network layers. Adam optimizer was used to optimize the AE network. The batch size, learning rate and max training epoch were set to be 64, 0.0001 and 500 respectively.

Moreover, 1000 samples were selected randomly from U as the training samples and the rest as test samples. For comparison purposes, three classifiers were considered, including logistic regression (LR), a shallow neural network known as MLP and random forest (RF) to learn the DR results in spectral domain and spatial domain respectively. By comparing the DR time, classification accuracy and the retention of data structure information, we discussed the adaptability of HSI data to DR methods.

3.3. Experimental result analysis
Table 1 and Table 2 provided the classification accuracy of different DR methods in spectral and spatial domain. As can be seen from Table 1, the classification accuracy after DR by PCA, KPCA-linear and SVD was significantly higher than that after nonlinear DR methods. In order to analyse the obvious difference between the DR effect in linear and nonlinear methods, we further compressed the low-dimensional embedded features into 2 dimensions by t-SNE and visualized the 2-dimensional features in the scatter diagram to the same coordinates range. The results were shown in Figure 2. Different colors of data points in the figure represented different categories of data points. As can be seen from the Figure 2 that after DR by PCA, KPCA-linear and SVD, the data points distribution had strong cluster effect, that is, data points of same category gathered together and great differences could be observed among data points of different categories. This phenomenon meant high data points separability and was helpful to improve the classification performance. Compared with KPCA-linear, data points distribution after KPCA-rbf was more confused, which led to the decline of classification performance. However, after DR by ICA, Isomap, LLE, LE and AE, data points of different categories were seriously confused which meant small differences and poor separability among them and this phenomenon directly led to the poor classification performance. ICA decomposed the original features into d independent variables. So, the setting of d was very important. In this experiment, d was set to 15, which was too small to separate data points of different categories. For manifold-based DR methods such as Isomap, LLE and LE, they focused on the local structure information among data points in the DR process but didn't consider the distribution trend of global data points. The training process of AE network only focused on the reconstruction error which also failed to estimate the global data points distribution trend and eventually led to the poor classification performance.

Table 2. Classification accuracy of different DR methods in spatial domain (%).

Classifier	Linear DR Methods				Nonlinear DR Methods				
	PCA	KPCA-linear	ICA	SVD	KPCA-rbf	Isomap	LLE	LE	AE
LR	80.9	**81.0**	**81.0**	80.9	72.0	53.2	48.8	55.8	58.1
MLP	90.3	90.3	**91.2**	90.4	81.9	59.7	51.7	57.9	71.1
RF	**89.4**	89.3	88.2	**89.4**	79.8	59.3	57.3	58.9	69.7

The results of spatial domain DR were shown in Table 2. The learning effects of linear DR methods were obviously better than that of nonlinear DR methods. Using t-SNE to visualize the distribution trend of low-dimensional embedded samples after spatial domain DR, it could be found the low-dimensional embedded samples obtained by linear DR had higher separability in space, which also showed that the linear structure feature was the main feature in Xiongan data set.

Table 3. Time-consuming of spectral and spatial domain DR (s).

Time Consuming (s)	Linear DR Methods				Nonlinear DR Methods				
	PCA	KPCA-linear	ICA	SVD	KPCA-rbf	Isomap	LLE	LE	AE
Spectral Domain	10.6	180.8	36.7	367.5	195.5	4248.7	3836.1	1653.5	15904.9
Spatial Domain	10.8	181.4	40.6	367.6	196.1	4250.7	3837.0	1657.9	15968.8

The runtime of various DR methods in spectral domain and spatial domain was shown in Table 3. Regardless of DR in spectral domain or spatial domain, Isomap, LLE and LE needed to calculate local structure among data points and AE needed to optimize the network parameters in iteration. So, these nonlinear DR methods had higher computational complexity and longer time-consuming than that of PCA, ICA and SVD in linear DR methods. Meanwhile, compared with KPCA-linear, the calculation complexity of KPCA employing radial basis kernel function was much higher. Therefore, KPCA-rbf took more time than KPCA-linear.

4. Conclusions

Linear and nonlinear DR methods were employed to reduce the dimension of samples selected from Xiongan data set in spectral domain and spatial domain respectively. By comparing the classification results, the linear DR methods could better preserve the intrinsic structure information of the original features and the samples of different categories had strong separability. That is, the linear structure features account for the main part in Xiongan data set. In addition, the time-consuming of linear DR methods was much less than that of nonlinear DR methods. The comparison of these two aspects showed that linear DR methods had better adaptability to Xiongan data set. Furthermore, it could be concluded that DR performance of PCA in spectral domain and spatial domain was the best. This paper also provided a new idea to measure the adaptability between the DR methods and the data set itself. In the future work, we will carry out DR experiments in more complex high-dimensional image data to explore the quantitative metrics of the adaptability between the DR methods and the data sets.

References

[1] Cao, F., Yang, Z., Hong, X., Cheng, Y., Huang, Y., & Lv, J. (2021). Supervised Dimensionality Reduction of Hyperspectral Imagery Via Local and Global Sparse Representation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 3860-3874.

[2] Chen, S., & Zhang, D. (2010). Semisupervised dimensionality reduction with pairwise constraints for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters, 8(2), 369-373.

[3] Balakrishnama, S., & Ganapathiraju, A. (1998). Linear discriminant analysis-a brief tutorial. Institute for Signal and information Processing, 18(1998), 1-8.

[4] Zabalza, J., Ren, J., Yang, M., Zhang, Y., Wang, J., Marshall, S., & Han, J. (2014). Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing, 93, 112-122.

[5] Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. Neural networks, 13(4-5), 411-430.

[6] Tenenbaum, J. B., De Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. science, 290(5500), 2319-2323.

[7] Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. science, 290(5500), 2323-2326.

[8] Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. Neural computation, 15(6), 1373-1396.

[9] Zabalza, J., Ren, J., Zheng, J., Zhao, H., Qing, C., Yang, Z., ... & Marshall, S. (2016). Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging. Neurocomputing, 185, 1-10.

[10] Cen Y, Zhang L F, Zhang X, Wang Y M, Qi W C, Tang S L, & Zhang P. (2020). Aerial hyperspectral remote sensing classification dataset of Xiongan New Area (Matiwan Village). Journal of Remote Sensing, 24(11), 1299-1306.

[11] Schölkopf, B., Smola, A., & Müller, K. R. (1997). Kernel principal component analysis. In International conference on artificial neural networks. Berlin. pp. 583-588.

[12] Kolda, T. G., & O'leary, D. P. (1998). A semidiscrete matrix decomposition for latent semantic indexing information retrieval. ACM Transactions on Information Systems (TOIS), 16(4), 322-346.

[13] Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(2605), 2579-2605.