Review on graph learning for dimensionality reduction of hyperspectral image

Liangpei Zhang & Fulin Luo

To cite this article: Liangpei Zhang & Fulin Luo (2020): Review on graph learning for dimensionality reduction of hyperspectral image, Geo-spatial Information Science, DOI: 10.1080/10095020.2020.1720529

To link to this article: https://doi.org/10.1080/10095020.2020.1720529
Review on graph learning for dimensionality reduction of hyperspectral image

Liangpei Zhang and Fulin Luo

State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

ABSTRACT

Graph learning is an effective manner to analyze the intrinsic properties of data. It has been widely used in the fields of dimensionality reduction and classification for data. In this paper, we focus on the graph learning-based dimensionality reduction for a hyperspectral image. Firstly, we review the development of graph learning and its application in a hyperspectral image. Then, we mainly discuss several representative graph methods including two manifold learning methods, two sparse graph learning methods, and two hypergraph learning methods. For manifold learning, we analyze neighborhood preserving embedding and locality preserving projections which are two classic manifold learning methods and can be transformed into the form of a graph. For sparse graph, we introduce sparsity preserving graph embedding and sparse graph-based discriminant analysis which can adaptively reveal data structure to construct a graph. For hypergraph learning, we review binary hypergraph and discriminant hyper-Laplacian projection which can represent the high-order relationship of data.

1. Introduction

Hyperspectral imaging technique captures the image data with a large number of consecutive narrow bands spanning the visible-to-infrared spectrums (Chen et al. 2019; Li et al. 2019; Zhang, Li, and Du 2019). Owing to differences in reflectivity for different materials under different electromagnetic spectrums, this technique is very effective to discriminate the composition of material and has been widely used in the fields of agriculture, forestry, geology, oceanography, meteorology, hydrology, military, environmental monitoring (Cao et al. 2019; Liu et al. 2017; Pan et al. 2019; Yakovliev et al. 2019; Chen, Xiao, and Li 2016). In these applications, a fundamental task is to obtain the class types in Hyper-Spectral Image (HSI) (Amini et al. 2018). Based on the imaging procedure, HSI contains rich spectral and spatial features (Zhang and Du 2012; Natsagdorj et al. 2017). For a large number of narrow bands, they have a strong correlation that results in massive redundant information in HSI (Du et al. 2018; Mohantly, Happy, and Routray 2019a). Moreover, the pixels in HSI are mixed because of the impact of imaging condition. In addition, the hundreds of bands make HSI data possess high dimensionality which will cause the Hughes phenomenon with the traditional classification methods (An et al. 2019; Mohantly, Happy, and Routray 2019b). For the Hughes phenomenon, the classification results will improve and then decrease with the increasing of dimensionality under limited samples (Hang and Liu 2018; Song et al. 2019; Zhao et al. 2017). Although HSI generates better discriminant performance for similar materials than the traditional image, it also faces more challenges to process HSI with the above problems.

To reduce the redundant information and enhance the discriminant power of features, an effective manner is to transform the original features into a new space that has lower dimensionality and stronger discriminant power (Deng et al. 2018; Li, Wang, and Cheng 2019; Huang et al. 2017; Frazier, Wang, and Chen 2014). With the transformed features, we can obtain better classification results. Principal Component Analysis (PCA) is a classic feature transformation method, which finds a set of orthogonal space to maximize the variance in each coordinate axis (Licciardi and Chanussot 2018; Liu, Singleton, and Arribas-Bel 2019). PCA has been widely used for HSI analysis to obtain the principal component features. Linear Discriminant Analysis (LDA) is a classic supervised method to obtain strong discriminant features of HSI (Peng and Luo 2016). LDA utilizes the priori class label to maximize the between-class variance and minimize the within-class variance which will separate the inter-class samples and compact the intra-class samples. Based on LDA, Nonparametric Weighted Feature Extraction (NWFE) is developed for dimensionality reduction in HSI classification task (Kuo, Li, and Yang 2009). NWFE sets different weights on each sample to calculate the weighted means and construct nonparametric between-class and within-class scatter matrices. To enhance anti-noise ability, Maximum Noise Fraction (MNF) was proposed by maximizing the signal-noise-ratio on the basis of the signal variance and the noise variance.

CONTACT Fulin Luo luoflyn@163.com

© 2020 Wuhan University. Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
In addition, Maximum Margin Criterion (MMC) was proposed with the marginal criteria to enhance the differences between different classes of HSI (Datta, Ghosh, and Ghosh 2017). These methods are based on the statistical properties of HSI, which neglects the intrinsic geometric structures.

To reveal the intrinsic structures of data, manifold learning was designed to discover the geometric properties of data. Three classic methods were generated including Isometric Mapping (ISOMAP) (Tenenbaum, de Silva, and Langford 2000), locally Linear Embedding (LLE) (Roweis and Saul 2000), and Laplacian Eigenmaps (LE) (Belkin and Niyogi 2003). In ISOMAP, the geodesic distance was used to represent the geometric relationships and it preserves the geodesic distances between samples in low-dimensional space. In LLE, it uses the local linear reconstruction to reveal the intrinsic structures of data and maintains the reconstruction relationships in low-dimensional space. In LE, it applies the local neighbor information to construct a Laplacian graph and does not change the structure of this graph in low-dimensional space. However, these manifold learning algorithms belong to a nonlinear projection which does not exist an explicit projection matrix to map the out-of-samples into the low-dimensional space. To address the out-of-sample problem, two linear projection methods were proposed on the basis of LLE and LE, which are called Neighborhood Preserving Embedding (NPE) (Huang and Huang 2014) and Locality Preserving Projections (LPP) (Zhai et al. 2016). NPE and LPP have the same theoretical basis as LLE and NPE, while they can obtain an explicit projection matrix to directly generate the low-dimensional features of out-of-sample. Based on the manifold learning methods, researchers have discovered that HSI exists manifold structure and introduced these methods into dimensionality reduction of HSI, which can reveal the intrinsic manifold of HSI.

On the basis of the above methods, a unified framework was proposed to understand these methods. This framework can be represented by a graph (Yan et al. 2007). In this framework, PCA, LDA, LLE, LE, NPE, and LLP can be redefined as different graph learning methods. Their main differences are the construction of similarity matrix and constraint matrix. These graph learning methods are very effective to reveal the intrinsic similar relationships of data which can reflect the homogeneity of data. They have been widely used to learn the dimensional features of HSI. But HSI contains complex intrinsic structures, the traditional graph learning methods do not effectively represent the structure relationships of HSI generally. To better represent the intrinsic properties of HSI, some advanced methods were developed based on this graph framework in recent years. Marginal Fisher Analysis (MFA) is a newly developed method under the graph framework (Huang et al. 2019). MFA constructs two graphs to represent the intrinsic properties of data. In the two graphs, one is a similar graph to compact the homogeneity of intraclass samples, and the other one is a penalty graph to separate the heterogeneity of interclass samples. In addition, some other methods have Local Fisher Discriminant Analysis (LFDA) (Wang, Ruan, and An 2016), Regularized Local Discriminant Embedding (RLDE) (Zhou, Peng, and Chen 2015), and Local Geometric Structure Fisher Analysis (LGSFA) (Luo et al. 2017a), which are both based on the graph learning framework. LFDA combines the ideas of LDA and LPP, and it separates the between-class samples as much as possible and simultaneously maintains the within-class local information. RLDE is based on MFA to add two regularized terms, which can preserve the data diversity and address the singularity with limited training samples. LGSFA considers the neighbor information of neighbor samples which is a benefit to obtain the homogeneity of data. For the advanced graph learning methods, they use the fixed neighbors to represent the local structures of data. However, a fixed neighbors may be unreasonable to represent the local structure of each sample.

To adaptively reveal the intrinsic structures of data, sparse representation was introduced into graph learning. Sparse representation is to linearly reconstruct a sample with an over-complete dictionary (Peng, Li, and Tang 2019). For the reconstructed coefficients, most of them are zeros and only a few of them are non-zeros which are termed sparse coefficients. According to the sparse coefficients, the relationship of coefficients can be used to adaptively represent the intrinsic properties of data (Wright et al. 2009). Then, a sparse graph can be constructed based on the sparse relationships of data. The sparse graph is robust to data noise and possesses datum-adaptive structures for each sample. Based on the sparse graph, many sparse graph embedding methods have been proposed to extract the low-dimensional features of data. Sparsity Preserving Projection (SPP) is a classic sparse dimensionality reduction method, which preserves the sparse properties of data in low-dimensional space (Qiao, Chen, and Tan 2010). In some applications, SPP is also termed as Sparsity Preserving Graph Embedding (SPGE) (Ly, Du, and Fowler 2014) and Sparse Neighborhood Preserving Embedding (Sparse NPE) (Cheng et al. 2010). In addition, Sparsity Preserving Analysis (SPA) was proposed to represent the relationships of data (Luo et al. 2017b). SPA utilizes the sparse coefficients to construct a sparse graph and maintains the structures of a sparse graph in a new space. The two sparse methods are both unsupervised which do not use any priori information in the dimensionality reduction models. To enhance the discriminant performance of models, some supervised graph methods have been developed based on sparse coefficients, including Discriminant Sparsity Neighborhood Preserving Embedding (DSNPE) (Lu, Jin, and Zou 2012).
Discriminative Learning by Sparse Representation Projections (DLSP) (Zang and Zhang 2011), Sparse Graph-based Discriminant Analysis (SGDA) (Ly, Du, and Fowler 2014) and Sparse Discriminant Embedding (SDE) (Huang and Yang 2015). For DSNPE, it incorporates sparse graph and MMC, which represents the within-neighboring information by the sparse reconstruction weights of the samples from the same class and the between-neighboring information by the sparse reconstruction weights of the neighboring samples from different classes. For DLSP, it combines local interclass geometrical structure and sparsity property, which can make the local within-class compactness. For SGA, it uses the class label information and sparse coefficients to construct an interclass reconstruction model and an intraclass reconstruction model, which can enhance the discriminant power of the model for HSI. For SDE, it utilizes the sparse reconstruction information and the interclass information to enhance the inter-manifold separability of data. For these graph learning methods, they just focus on the binary relationships of data that only two samples are connected with an edge. However, HSI contains complex high-order relationships. Therefore, the traditional graph learning methods cannot effectively reveal the intrinsic structures of HSI.

To represent the complex high-order structures, hypergraph was introduced into machine learning and many hypergraph learning methods were generated to better reveal intrinsic structures of data (Ji et al. 2014). Binary Hypergraph (BH) is a classic method for the feature representation of HSI which uses k nearest neighbors to construct the hypergraph model (Yuan and Tang 2015). To utilize the class label information, a supervised hypergraph was proposed to improve the compactness of low-dimensional features, which is termed as Discriminant Hyper-Laplacian Projection (DHLP) (Huang et al. 2016). With the spatial information, a new hypergraph learning method was generated to analyze the spatial-spectral structures of HSI, called Hyper-Graph embedding-based Spatial-Spectral joint features (SSHG) (Sun et al. 2017). However, SSHG just uses the stack manner to represent the spatial-spectral features, which will not enough extract the intrinsic features. Therefore, the Spatial-Spectral Hyper-Graph Discriminant Analysis (SSHGDA) method was proposed to better reveal the spatial-spectral features and enhance the discriminating power of low-dimensional features (Luo et al. 2019). In addition, the class label is very difficult to obtain in real application. Based on this, the Semi-supervised Hyper-Graph Embedding (SHGE) method was developed to construct a feature learning model with the label and unlable samples (Du et al. 2018).

To understand the development of graph learning method, we review several classic and show some experiments to analyze these methods in this paper. In Section 2, we introduce two basic theories including graph and hypergraph. Section 3 reviews several graph learning methods. Some experiments are shown in Section 4. Finally, Section 5 provides some conclusions.

2. Basic theory

In this paper, we represent a pixel of HSI as a vector $x_i \in \mathbb{R}^D (i = 1, 2, \ldots, n)$, where $n$ and $D$ are the number of pixels and bands, respectively. $\ell(x_i) \in \{1, 2, \cdots, c\}$ denotes the class label of $x_i$, where $c$ is the class number of land-cover. A HSI data set is denoted as $X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{D \times n}$. $Y = [y_1, y_2, \ldots, y_n] \in \mathbb{R}^{d \times n}$ represents the low-dimensional features of $X$, where $d$ is the embedding dimension. With a projection matrix $V \in \mathbb{R}^{D \times d}$, the low-dimension features can be represented as $Y = V^TX$.

2.1. Graph learning

Graph is used to reflect the relationship of two samples, which can represent some of the statistical or geometrical properties of data. For a graph $G$, it can be denoted as $G = \{X, E, W\}$ that is an undirected graph, where $X$ denotes the vertices, $E$ denotes the edges, and $W = [w_{ij}]_{i,j=1}^{n}$ represents the weight matrix of edges. To construct a graph, the neighbors are connected by edges and a weight is given to the corresponding edges. If vertices $i$ and $j$ are similar, we should connect an edge between vertices $i$ and $j$, and a similar weight is also defined for the edge.

In low-dimensional space, we should preserve the structures of graph. Therefore, the objective function can be defined as

$$\min \frac{1}{2} \sum_{i,j} \left\| y_i - y_j \right\|^2 w_{ij} = \min \text{tr}(LY^T)$$

subject to $\text{tr}(YHY^T) = h$

where $h$ is a constant, $H$ is a constraint matrix to avoid the trivial solution, $L = D - W$ is the Laplacian matrix of graph $G$, $D = \text{diag}([\sum_{j=1}^{n} w_{ij}, \sum_{j=1}^{n} w_{2j}, \ldots, \sum_{j=1}^{n} w_{nj}])$ is a diagonal matrix. In general, $H$ can be an identity matrix for scale normalization or a Laplacian matrix of penalty graph to restrain some unwanted similarity.

2.2. Hypergraph learning

For hypergraph and graph, the difference is the number of vertex on each edge. In graph, each edge is only composed of two vertices, while each edge of hypergraph has more than two vertices. For a hypergraph, it is defined as $G^H = \{X, E^H, W^H\}$, where $X$ is the vertices, $E^H$ is the hyperedges, $W^H$ is the weights, and each hyperedge $e_i \in E^H$ has a weight $w_i \in H = \{h_{ij}\}_{i,j \in}$
$H^{[v_i \times E^H]}$ is an incidence matrix to reveal the relationship between vertex $v_i$ and hyperedge $e_j$, and $h^{v_i}_{e_j}$ is
\[
h^{v_i}_{e_j} = \begin{cases} 1, & \text{if } v_i \in e_j \\ 0, & \text{otherwise} \end{cases}
\] (2)

For the hypergraph, the degrees of vertex $v_i$ and hyperedge $e_j$ are used to represent the properties of hypergraph, which are the summation of weights via $v_i$ and the number of vertices on $e_j$. Therefore, the degrees of $v_i$ and $e_j$ are
\[
d^v_i = \sum_{j=1}^{n} w^v_{ij} h^{v_i}_{e_j} \\
d^e_j = |e_j| = \sum_{i=1}^{n} h^{v_i}_{e_j}
\] (3)

In low-dimensional space, the structures of hyperedge should be preserved to reveal the high-order properties of HSI, and the embedding function is
\[
\min_{Y} \frac{1}{2} \sum_{e_i \in E^H} \sum_{(x_k, x_j) \in e_i} \left\| y_j - y_k \right\|^2 \\
= \min_{Y} \text{tr}(YL^H Y^T) \\
s.t. \text{tr}(\frac{1}{n}YD^vY^T) = 1
\] (4)

where $L^H = D^v - HW^H (D^v)^{-1} (H)^T$ is the hyper-Laplacian matrix. $W^H = \text{diag}([w^v_{11}, w^v_{22}, \ldots, w^v_{n_n}])$ is the weight matrix. $D^v = \text{diag}([d^v_1, d^v_2, \ldots, d^v_n])$ and $D^e = \text{diag}([d^e_1, d^e_2, \ldots, d^e_n])$ are the diagonal matrices to denote the degrees of vertex and hyperedge.

Figure 1 explains the difference between graph and hypergraph. In Figure 1(a), each edge only has two vertices for the graph (i.e. $v_1$ and $v_2$). In Figure 1(b), this hypergraph contains seven vertices and four hyperedges, and each hyperedge contains more than two vertices (i.e. $v_1$, $v_2$, and $v_4$ belonging to $e_1$).

### 3. Representative graph learning methods

In this section, we review several graph learning methods including LPP, NPE, SPGE, SGDA, BH, and DHLP.

#### 3.1. Locality preserving projections

LPP is used to learn the intrinsic manifold structures of high-dimensional data. In LPP, a similar graph is designed with $K$ nearest neighbors. Then, it preserves the neighbor structures in low-dimensional space, that is to say, the graph structure is maintained in feature embedding space.

According to the neighbor relationship of each sample, a similar weight between two samples is defined by the thermonuclear function ($t$ is thermonuclear parameter).
\[
w_{ij} = \begin{cases} \exp(-\|x_i - x_j\|^2), & \text{if } x_j \text{ is the neighbors of } x_i \\ 0, & \text{otherwise} \end{cases}
\] (5)

where $w_{ij}$ is the weight between $x_i$ and $x_j$. The weight matrix is $W = [w_{ij}]_{i,j}$.

In low-dimensional space, the neighbor structure of each sample is preserved and the similar samples should be compacted as much as possible. The minimization objective function is defined as
\[
\min_{V} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left\| V^T x_i - V^T x_j \right\|^2 w_{ij} = \min_{V} \text{tr}(V^T XLX^T V) \\
s.t. \text{tr}(V^T XDX^T V) = I
\] (6)

where $L = D - W$ is the Laplacian matrix. $D = \text{diag}([d_{11}, d_{22}, \ldots, d_{nn}])$ is a diagonal matrix and $d_{ii} = \sum_{j=1}^{n} w_{ij}$.

This optimization function can be equal to the following generalized eigenvalue problem.
\[
XLX^T v_i = \lambda_i XDX^T v_i
\] (7)
Then, the projection matrix is composed of the $d$ smallest eigenvalues corresponding eigenvectors.

$$V = [v_1, v_2, \cdots, v_d] \in \mathbb{R}^{D \times d}$$ (8)

### 3.2. Neighborhood preserving embedding

NPE assumes that data possesses local linear structure. In low-dimensional space, the local linear structure should be maintained to reveal the intrinsic manifold properties of data. In LPP, it uses $K$ nearest neighbors to represent the local information.

According to the local structure, the reconstruction error is used to calculate the reconstruction weights as follows:

$$\min_{w} \sum_{i=1}^{n} ||x_i - \sum_{j=1}^{n} w_{ij} x_j||^2$$

where $w_{ij}$ is the weight between $x_i$ and $x_j$, $\sum_{j=1}^{n} w_{ij} = 1$, if $x_j$ is the neighbor of $x_i$, $w_{ij} \neq 0$, otherwise $w_{ij} = 0$.

To solve this function, we can obtain the weight as

$$w_{ij} = \frac{\sum_{m=1}^{k} (z_{jm})^{-1} z_{ij}}{\sum_{j=1}^{k} \sum_{m=1}^{k} (z_{jm})^{-1}}$$ (9)

where $z_{mm} = (x_i - x_m)^T (x_i - x_m)$, $z_i = [x_i - x_1, x_i - x_2, \cdots, x_i - x_k]^T$, $z_{ij} = [z_{mm}]_{k \times k}$, $x^j_i$ is the $j$-th neighbor of $x_i$.

In low-dimensional space, the weights are preserved to extract the embedding features, and the objective function is defined as

$$\min_{V} \sum_{i=1}^{n} \left\| V^T x_i - \sum_{j=1}^{n} w_{ij} V^T x_j \right\|^2 = \text{tr}(V^T X M X^T V)$$

s.t. $\frac{1}{n} V^T X X^T V = I$ (10)

where $M = (I - W)^T (I - W)$.

For the optimization function, it can be transformed to a generalized eigenvalue problem.

$$X M X^T v = \lambda X X^T v$$ (11)

According to the $d$ smallest eigenvalues corresponding eigenvectors. The projection matrix can be denoted as

$$V = [v_1, v_2, \cdots, v_d] \in \mathbb{R}^{D \times d}$$ (12)

### 3.3. Sparsity preserving graph embedding

SPGE utilizes sparse coefficients to construct a sparse graph that reveals the sparse properties of data. It preserves the sparse graph structures in low-dimensional space. SPGE inherits the natural discriminating power of sparse representation to represent the intrinsic relationships of data.

For a sample $x_i$, it can be represented by a dictionary $X_i$ where $X_i = [x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n]$, and the representation coefficients are very sparse. The optimization function is denoted as

$$\min_{s_i} ||s_i||_1$$

s.t. $||x_i - W_is_i||^2_F < \epsilon$ (13)

where $\epsilon > 0$ is error tolerance, $s_i = [s_{i1}, s_{i2}, \cdots, s_{i,j}, \cdots, s_{i,n}]$ is the sparse coefficients, $||s_i||_1$ denotes the $\ell_1$-norm of $s_i$ that is used to control the sparsity of $s_i$.

According to the sparse coefficients $s_i$, a graph $G = \{X, W_s\}$ is constructed, where $W_s$ is the similar matrix to reflect the similarity between two vertices, and it is defined as

$$(W_s)_{m,n} = \begin{cases} 0, & m = m' \cr s_{m,n}, & m > m' \cr s_{m,n} - 1, & m > m' \end{cases}$$ (14)

Once the sparse graph is obtained, a dimensionality reduction function can be constructed to minimize the low-dimensional reconstruction error for the sparse representation, an optimization function is formulated as

$$\min_{V} \sum_{i=1}^{N} \left\| V^T x_i - \sum_{m=1}^{m'} (W_s)_{m,n} V^T X_{nm} \right\|^2 = \text{tr}(V^T X L_s X^T V)$$

s.t. $\frac{1}{N} V^T X X^T V = I$ (15)

where $L_s = (I - W_s)^T (I - W_s)$, $V$ is the transformation matrix.

To avoid degenerate solutions, a constraint $V^T X X^T V = I$ is added.

$$\min_{V} \text{tr}(V^T X L_s X^T V)$$

s.t. $\frac{1}{N} V^T X X^T V = I$ (16)

This optimization problem can be represented as a generalized eigenvalue problem.

$$X L_s X^T v = \lambda X X^T v$$ (17)

With the $d$ smallest eigenvalues corresponding eigenvectors, the projection matrix can be obtained as

$$V = [v_1, v_2, \cdots, v_d] \in \mathbb{R}^{D \times d}$$ (18)

### 3.4. Sparse graph-based discriminant analysis

SPGE is an unsupervised method that restrains the discriminant power of feature representation. To make full use of the priori information, a supervised method was formulated with the class label information based on SPGE, termed sparse graph based discriminate analysis (SGDA). SGDA utilizes the sparse
coefficients and the class label information to define a with-class reconstruction and a between-class reconstruction, which can improve the intraclass compactness and interclass separability.

For a sample \( x_m \) from the \( z \)-th class, a selector function is defined as
\[
S_z(x_m) = \begin{cases} 1, & \text{if } \ell(x_m) = z \\ 0, & \text{otherwise} \end{cases}
\]  
(19)

According to the sparse coefficients from (13), a similarity matrix \( W_z \) can be obtained by (14). Then, a within-class reconstruction error is defined as
\[
S_w = \sum_{z=1}^{c} \sum_{m=1}^{n} S_z(x_m) \left\| V^T x_m - \sum_{m'=1}^{n} S_z(x_{m'}) (W_z)_{m,m'} V^T x_{m'} \right\|^2
\]  
(20)

If we define an intraclass similarity matrix \( W_z' \) as
\[
(W_z')_{m,m'} = \begin{cases} (W_z)_{m,m'}, & \ell(x_m) = \ell(x_{m'}) \\ 0, & \text{otherwise} \end{cases}
\]  
(21)

then the objection can be simplified as
\[
S_w = \text{tr}(V^T X(I - W_z')^T(I - W_z')X^TV)
\]  
(22)

At the same time, for the interclass samples, a between-class reconstruction error is denoted as
\[
S_b = \sum_{z=1}^{c} \sum_{m=1}^{n} S_z(x_m) \left\| V^T x_m - \sum_{m'=1}^{n} (1 - S_z(x_{m'}))(W_z)_{m,m'} V^T x_{m'} \right\|^2
\]  
(23)

which can transformed as
\[
S_b = \text{tr}(V^T X(I - W_z')^T(I - W_z')X^TV)
\]  
(24)

where
\[
(W_z')_{m,m'} = \begin{cases} (W_z)_{m,m'}, & \ell(x_m) \neq \ell(x_{m'}) \\ 0, & \text{otherwise} \end{cases}
\]  
(25)

In low-dimensional space, an optimization function is constructed to minimize the within-class reconstruction error while maximizing the between-class reconstruction error, i.e.
\[
\min_{V} \frac{S_w}{S_b} = \min_{V} \frac{\text{tr}(V^T XLX^TV)}{\text{tr}(V^T L_p X^TV)}
\]  
(26)

where
\[
L_z = (I - W_z')^T(I - W_z')
\]  
(27)
\[
L_p = (I - W_z')^T(I - W_z')
\]  
(28)

For the optimization problem, it can be represented as a generalized eigenvalue problem.
\[
XLX^TV_l = \lambda_l XL_p X^TV_l
\]  
(29)

Then, we can obtain the \( d \) smallest eigenvalues corresponding eigenvectors to compose the projection matrix as
\[
V = [v_1, v_2, \ldots, v_d] \in \mathbb{R}^{D \times d}
\]  
(30)

### 3.5. Binary hypergraph

BH can reveal the high-order relationship of data, which is more effective for real applications. BH uses \( K \) nearest neighbors to construct the hypergraph in general. For a hyperedge \( e_j \), it is composed of a centroid vertex \( x_i \) and its \( K \) nearest neighbor; thus, each hyperedge has \( K + 1 \) pixels.

According to the hypergraph, the incidence matrix \( H \) can be represented as
\[
H_{ij} = h(x_i, e_j) = \begin{cases} 1, & \text{if } x_i \text{ is the neighbor of } x_j \\ 0, & \text{otherwise} \end{cases}
\]  
(31)

The hyperedge weight \( w_i \) is defined as the following.
\[
w_i = w(e_i) = \sum_{x_i \in e_i} \exp(-\frac{||x_j - x_i||^2}{t})
\]  
(32)

where \( t \) is a parameter.

According to the incidence matrix and the hyperedge weights, the degrees of each vertex and each hyperedge are represented as
\[
d_i = d(x_i) = \sum_{j=1}^{n} w_{ij} H_{ij}
\]  
(33)
\[
\delta_i = d(e_i) = \sum_{j=1}^{n} H_{ji}
\]  
(34)

With the weights, the vertex degree, and the hyperedge degree, a feature embedding objective function is designed as
\[
\min_{V} \frac{1}{2} \sum_{e_j \in E} \sum_{(x_i, x_k) \in e_j} \frac{w(e_i) h(x_j, e_j) h(x_k, e_k)}{d(e_i)} \left\| V^T x_j - V^T x_k \right\|^2
\]  
\[+ \frac{1}{2} \sum_{t} \sum_{j \in \delta_t} \frac{w_{ij} H_{ij}}{d_i} \left\| V^T x_j - V^T x_t \right\|^2
\]  
(35)

where \( L^{BH} = D^* - HW^{BH}(D^*)^{-1}H^T \) is the hyper-Laplacian matrix. \( W^{BH} = \text{diag}([w_1, w_2, \ldots, w_n]) \) is the weight matrix. \( D^* = \text{diag}([d_1, d_2, \ldots, d_n]) \) and \( D^* = \text{diag}([\delta_1, \delta_2, \ldots, \delta_n]) \) are the diagonal matrices to denote the degrees of vertex and hyperedge.

To obtain a unique solution, a constrain \( \text{tr}(V^T XD^* X^TV) = 1 \) is added. The optimization solution is transformed to a generalized eigenvalue problem.
\[ X \lambda X^T v_i = \lambda X D^T X^T v_i \]  
(36)

The projection matrix can be obtained by the \( d \) smallest eigenvalues corresponding eigenvectors.

\[ V = [v_1, v_2, \cdots, v_d] \in \mathbb{R}^{D \times d} \]  
(37)

### 3.6. Discriminant hyper-Laplacian projection

To enhance the discriminating power of BH, a supervised method is proposed with the class label information. According to hypergraph, a normalization hypergraph learning model is constructed as

\[
\min_{V} \frac{1}{2} \sum_{i \in E} \sum_{(x_j, x_k) \in E_i} \frac{w(e_i) h(x_j, e_i) h(x_k, e_i)}{d(e_i)} \left| V^T x_j - V^T x_k \right|^2 \left| d(x_j) - d(x_k) \right| 
\]

\[
= \text{tr}(V^T X L^{NH} X^T V) 
\]

(38)

where \( L^{NH} = I - (D^T)^{-1/2} HW^{NH} (D^T)^{-1/2} \) is the hyper-Laplacian matrix.

To improve the difference of intra-class samples, a maximizing objective function is defined as

\[
\max_{V} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{c} b_{ij} \left| V^T u_i - V^T u_j \right|^2 
\]

\[
= \text{tr}(V^T U L^U U^T V) 
\]

(39)

where \( U = [u_1, u_2, \cdots, u_c] \) is the mean matrix, \( u_i \) is the \( i \)-th class mean. \( L^U = D^u - B \) is the Laplacian matrix of classes. \( D^u = \text{diag}([\sum_j b_{ij}, \sum_j b_{ij}, \cdots, \sum_j b_{ij}]) \) is a diagonal matrix. The mean similar weight matrix \( B = [b_{ij}] \) is defined as

\[
b_{ij} = \exp\left(-\frac{\left\| u_i - u_j \right\|^2}{t}\right) 
\]

(40)

Therefore, the objective function of DHLP is represented as

\[
\min_{V} \frac{\text{tr}(V^T X L^{NH} X^T V)}{\text{tr}(V^T U L^U U^T V)} 
\]

(41)

For the optimization function, it can be equal to a generalized eigenvalue problem.

\[ X L^{NH} X^T v_i = \lambda U L^U U^T v_i \]  
(42)

According to the \( d \) smallest eigenvalues corresponding eigenvectors, we can get

\[ V = [v_1, v_2, \cdots, v_d] \in \mathbb{R}^{D \times d} \]  
(43)

### 4. Conclusions

Hyperspectral image possesses hundreds of narrow bands, which exists massive redundant information and Hughes phenomenon for the traditional classification methods. To address the challenges, graph learning is an effective technique to represent the intrinsic properties of data, which can be used to reduce data redundancy and dimensionality. In this paper, we review several graph learning methods for dimensionality reduction of hyperspectral image. The reviewed methods include two classic manifold methods (i.e. LPP and NPE, and they can be transformed into a graph framework), two sparse graph methods (i.e. SPGE and SGDA, and they can adaptively reveal the intrinsic relationship of data), and two hypergraph learning methods (i.e. BH and DHLP, and they can represent the high-order structures of data). According to the review, we will better understand the graph learning method and develop some new methods. In this paper, we just review several basic algorithms based on the spectral information. Therefore, we can use the spatial-spectral information to construct spatial-spectral graph learning in the future.

### Acknowledgments

We would like to thank the handling editor and the anonymous reviewers for their detailed and constructive comments and suggestions, which indeed helped to improve the quality of this paper.

### Disclosure statement

No potential conflict of interest was reported by the authors.

### Funding

This work is supported by the National Natural Science Foundation of China [grant number 61801336], the China Postdoctoral Science Foundation [grant number 2019M662717 and 2017M622521], and the China Postdoctoral Program for Innovative Talent [grant number BX201700182].

### Notes on contributors

**Liangpei Zhang** received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi’an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi’an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998. He is currently the Head of the Remote Sensing Division with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He is also a “Chang-Jiang Scholar” Chair Professor appointed by the Ministry of Education of China, Beijing, China, and a Principal Scientist for the China State Key Basic Research project 2011-2016 appointed by the Ministry of the National Science and Technology of China to lead the remote sensing program in China. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

**Fulin Luo** received the Ph.D and M.S. degree in Instrument Science and Technology from Chongqing University,
Chongqing, China, in 2016 and 2013, respectively. He is currently an Associate Researcher with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University. His research interests are hyperspectral image classification, image processing, sparse representation and manifold learning in general.

ORCID

Liangpei Zhang http://orcid.org/0000-0001-6890-3650
Fulin Luo http://orcid.org/0000-0002-7696-0775

References

Amini, S., S. Homayouni, A. Safari, and A. Darvishsefat. 2018. “Object-based Classification of Hyperspectral Data Using Random Forest Algorithm.” Geo-spatial Information Science 21 (2): 127–138. doi:10.1080/10095020.2017.1399674.

An, J. L., J. H. Lei, Y. Z. Song, X. R. Zhang, and J. M. Guo. 2019. “Tensor Based Multiscale Low Rank Decomposition for Hyperspectral Images Dimensionality Reduction.” Remote Sensing 11 (12): 1485. doi:10.3390/rs11121485.

Belkin, M., and P. Niyogi. 2003. “Laplacian Eigenmaps for Dimensionality Reduction and Data Representation.” Neural Computation 15 (6): 1373–1396. doi:10.1162/089976603321780317.

Cao, F. X., Z. J. Yang, J. C. Ren, W. Z. Chen, G. J. Han, and Y. Z. Shen. 2019. “Local Block Multilayer Sparse Extreme Learning Machine for Effective Feature Extraction and Classification of Hyperspectral Images.” IEEE Transactions on Geoscience and Remote Sensing 57 (8): 5580–5594. doi:10.1109/tgrs.2019.2900509.

Chen, F., D. Xiao, and Z. Li. 2016. “Developing Water Quality Retrieval Models with in Situ Hyperspectral Data in Poyang Lake, China.” Geo-spatial Information Science 19 (4): 255–266. doi:10.1080/10095020.2016.1258201.

Chen, Y., K. Zhu, L. Zhu, X. He, P. Ghamisi, and J. A. Benediktsson, 2019. “Automatic Design of Convolutional Neural Network for Hyperspectral Image Classification.” IEEE Transactions on Geoscience and Remote Sensing 57 (9): 7048–7066. doi:10.1109/tgrs.2019.2910603.

Cheng, B., J. C. Yang, S. C. Yan, Y. Fu, and T. S. Huang. 2010. “Learning with I(1)-Graph for Image Analysis.” IEEE Transactions on Image Processing 19 (4): 858–866. doi:10.1109/tip.2009.2038764.

Datta, A., S. Ghosh, and A. Ghosh. 2017. “Supervised Feature Extraction of Hyperspectral Images Using Partitioned Maximum Margin Criterion.” IEEE Geoscience and Remote Sensing Letters 14 (1): 82–86. doi:10.1109/lgrs.2016.2628078.

Deng, Y. J., H. C. Li, K. Fu, Q. Du, and W. J. Emery. 2018. “Tensor Low-Rank Discriminant Embedding for Hyperspectral Image Dimensionality Reduction.” IEEE Transactions on Geoscience and Remote Sensing 56 (12): 7183–7194. doi:10.1109/tgrs.2018.2849085.

Du, W. B., W. W. Qiang, M. Li, Q. L. Hou, L. Zhen, and L. Jing. 2018. “Semi-supervised Dimension Reduction Based on Hypergraph Embedding for Hyperspectral Images.” International Journal of Remote Sensing 39 (6): 1696–1712. doi:10.1080/01431161.2017.1415480.

Frazier, A., L. Wang, and J. Chen. 2014. “Two New Hyperspectral Indices for Comparing Vegetation Chlorophyll Content.” Geo-spatial Information Science 17 (1): 17–25. doi:10.1080/10095020.2014.889264.

Hang, R. L., and Q. S. Liu. 2018. “Dimensionality Reduction of Hyperspectral Image Using Spatial Regularized Local Graph Discriminant Embedding.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11 (9): 3262–3271. doi:10.1109/jstars.2018.2847042.

Huang, H., and Y. B. Huang. 2014. “Improved Discriminant Sparsity Neighborhood Preserving Embedding for Hyperspectral Image Classification.” Neurocomputing 136: 224–234. doi:10.1016/j.neucom.2014.01.010.

Huang, H., and M. Yang. 2015. “Dimensionality Reduction of Hyperspectral Images with Sparse Discriminant Embedding.” IEEE Transactions on Geoscience and Remote Sensing 53 (9): 5160–5169. doi:10.1109/tgrs.2015.2418203.

Huang, S., D. Yang, Y. X. Ge, and X. H. Zhang. 2016. “Discriminant Hyper-Laplacian Projections and Its Scalable Extension for Dimensionality Reduction.” Neurocomputing 173: 145–153. doi:10.1016/j.neucom.2015.01.010.

Huang, X., J. Li, W. Liao, and J. Chanussot. 2017. “Information Extraction from Remote Sensing Imagery.” Geo-spatial Information Science 20 (4): 297–298. doi:10.1080/10095020.2017.1404273.

Huang, Z. Y., H. Y. Zhu, J. T. Zhou, and X. Peng. 2019. “Multiple Marginal Fisher Analysis.” IEEE Transactions on Industrial Electronics 66 (12): 9798–9807. doi:10.1109/tie.2018.2870413.

Ji, R. R., Y. Gao, R. C. Hong, Q. Liu, D. C. Tao, and X. L. Li. 2014. “Spectral-Spatial Constraint Hyperspectral Image Classification.” IEEE Transactions on Geoscience and Remote Sensing 52 (3): 1811–1824. doi:10.1109/tgrs.2013.2255297.

Kuo, B. C., C. H. Li, and J. M. Yang. 2009. “Kernel Nonparametric Weighted Feature Extraction for Hyperspectral Image Classification.” IEEE Transactions on Geoscience and Remote Sensing 47 (4): 1139–1155. doi:10.1109/tgrs.2008.2008308.

Li, D. Q., X. S. Wang, and Y. H. Cheng. 2019. “Spatiospectral Neighbour Graph for Dimensionality Reduction of Hyperspectral Image Classification.” International Journal of Remote Sensing 40 (11): 4361–4383. doi:10.1080/01431161.2018.1562587.

Li, S. T., W. W. Song, L. Y. Fang, Y. S. Chen, P. Ghamisi, and J. A. Benediktsson. 2019. “Deep Learning for Hyperspectral Image Classification: An Overview.” IEEE Transactions on Geoscience and Remote Sensing 57 (9): 6690–6709. doi:10.1109/tgrs.2019.2907932.

Liccari, G., and J. Chanussot. 2018. “Spectral Transformation Based on Nonlinear Principal Component Analysis for Dimensionality Reduction of Hyperspectral Images.” European Journal of Remote Sensing 51 (1): 375–390. doi:10.1080/22797258.2018.1441670.

Liu, Y., A. Singleton, and D. Arribas-Bel. 2019. “A Principal Component Analysis (Pca)-based Framework for Automated Variable Selection in Geodemographic Classification.” Geo-spatial Information Science 22 (4): 251–264. doi:10.1080/10095020.2019.1621549.

Liu, Z., B. Tang, X. He, Q. Qiu, and F. Liu. 2017. “Class-Specific Random Forest with Cross-Correlation Constraints for Spectral-Spatial Hyperspectral Image Classification.” IEEE Geoscience and Remote Sensing Letters 14 (2): 257–261. doi:10.1109/lgrs.2016.2637561.

Lu, G., Z. Jin, and J. Zou. 2012. “Face Recognition Using Discriminant Sparsity Neighborhood Preserving
