SPATIAL-SPECTRAL HYPERSPECTRAL IMAGE CLASSIFICATION VIA MULTIPLE RANDOM ANCHOR GRAPHS ENSEMBLE LEARNING

Yanling Miao, Qi Wang, Mulin Chen, Xuelong Li*

School of Computer Science and School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P.R. China.

ABSTRACT

Graph-based semi-supervised learning methods, which deal well with the situation of limited labeled data, have shown dominant performance in practical applications. However, the high dimensionality of hyperspectral images (HSI) makes it hard to construct the pairwise adjacent graph. Besides, the fine spatial features that help improve the discriminability of the model are often overlooked. To handle the problems, this paper proposes a novel spatial-spectral HSI classification method via multiple random anchor graphs ensemble learning (RAGE). Firstly, the local binary pattern is adopted to extract the more descriptive features on each selected band, which preserves local structures and subtle changes of a region. Secondly, the adaptive neighbors assignment is introduced in the construction of anchor graph, to reduce the computational complexity. Finally, an ensemble model is built by utilizing multiple anchor graphs, such that the diversity of HSI is learned. Extensive experiments show that RAGE is competitive against the state-of-the-art approaches.

Index Terms— Hypersectral images, semi-supervised learning, anchor graph, spatial-spectral information, ensemble learning

1. INTRODUCTION

Hyperspectral images (HSI) obtained by hyperspectral imaging spectrometer provides abundant spatial structure and spectral information of the observed objects. With very narrow diagnostic spectral bands, HSI can effectively reflect subtle objects between land cover classes [1]. Therefore, HSI has been used in the real-world applications. Among these, the classification is considered as a fundamental task.

A plenty of conventional supervised methods, which only learn from labels, have been developed for HSI classification [2, 3]. Since the data labeling is quite costly, the models (i.e., semi-supervised learning) that utilize labeled and unlabeled data are rather meaningful. Due to the simple and elegant formulation, the semi-supervised learning (SSL) has obtained more attention [4]. Especially, graph-based SSL models belong to convex optimization problem and provide the

closed-form solution. However, graph-based SSL usually suffer from the high time complexity with computing the inverse of graph Laplacian [5]. To address the above problem, the anchor graph (AG)-based approaches have been developed, which exploit a small subset of the whole points defined anchors to construct the large-scale adjacent graph [6]. It reduces the computational complexity and effectively processes HSI data. Nevertheless, the previous AG-based works rarely use the spatial features within pixels, which limits their discriminability.

Many attempts have been made in spatial-spectral combined approaches [7, 8, 9]. The intraclass spatial-spectral hypergraph was constructed by considering the coordinate relationship and similarity between adjacent samples [10]. These methods usually extract the spatial features based on the coordinate distance metric between neighboring pixels, which fail to discovery the fine feature (i.e., local texture feature). In addition, the aforementioned models are based on a single learner, which are weak in diversity learning. To alleviate this situation, the ensemble learning strategy obtains more attention [11]. However, the AG-based ensemble learning method is rarely considered in HSI classification task.

To overcome the aforesaid shortcomings, this paper proposes a novel multiple Random Anchor Graphs Ensemble learning method (RAGE) for HSI classification. The main contributions can be summarized as follows.

• To capture the fine feature, we adopt the Local Binary Pattern (LBP) method to extract the texture information as the spatial feature, which can refine the local structures of HSI.

• To effectively handle the high-dimensional and large-scale problem, we propose an anchor graph-based SSL model, which learns the adjacent graph by assigning neighbors adaptively.

• To obtain the optimal predictive model, we introduce the ensemble learning strategy in the proposed scheme, where multiple anchor graphs are constructed in parallel to ensure the runtime efficiency.

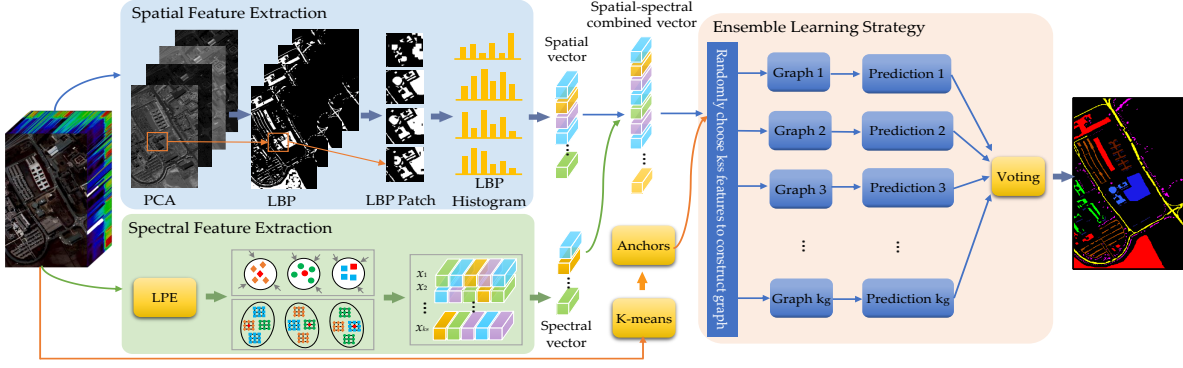*Qi Wang is the corresponding author.

Fig. 1. The overview architecture of the proposed RAGE.

2. METHODOLOGY

2.1. Graph-based Semi-supervised Learning Framework

Graph-based SSL methods utilize the undirected graph to learn the intrinsic geometric structures of HSI data and encode pairwise relations between pixels.

Suppose that each pixel of HSI can be denoted as $x_i \in \mathbb{R}^d$ $(i = 1, 2, ..., n)$, where d is the number of spectral bands and n refers to the number of HSI pixels. The HSI data is denoted as $\mathbf{X} = [x_1, x_2, ..., x_n]^T \in \mathbb{R}^{n \times d}$. Let $\mathbf{X} = \mathbf{X}_l \cup \mathbf{X}_v (n = l + v)$, where $\mathbf{X}_l = \{x_1, x_2, ..., x_l\}$ is the labeled set with the labels $\mathbf{Y}(x_i) \in \{1, 2, ..., c\}$ and $\mathbf{X}_v = \{x_l, x_{l+1}, ..., x_{l+v}\}$ is the unlabeled set. The s_{ij} represents the similarity between x_i and x_j, $\mathbf{S} \in \mathbb{R}^{n \times n}$. According to the manifold assumption, the adjacent points x_i and x_j are likely to be divided into the same category if the weight s_{ij} is large.

For the multi-class classification task, the graph-based SSL methods can be considered as the quadratic optimization problem:

$$\min_{\mathbf{F}} \operatorname{Tr}((\mathbf{F} - \mathbf{Y})^{\mathbf{T}} \mathbf{C}(\mathbf{F} - \mathbf{Y})) + \operatorname{Tr}(\mathbf{F}^{\mathbf{T}} \mathbf{L} \mathbf{F}), \quad (1)$$

where $\mathbf{F} \in \mathbb{R}^{n \times c}$ is the predicted labels matrix. $\mathbf{Y} = [y_1, ..., y_l, 0, ..., 0]^T \in \mathbb{R}^{n \times c}$ denotes the labels matrix. $\mathbf{C} \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose i-th element is represented as $c_i = \alpha_l$ $(1 \le i \le l)$ and $c_i = \alpha_v$ $(l + 1 < i \le n)$, where α_l and α_v are two parameters. $\mathbf{L}$ is the Laplacian matrix, where $\mathbf{L} = \mathbf{D} - \mathbf{S}$ and $\mathbf{D}$ is the degree matrix.

2.2. Spatial and Spectral Feature Extraction

The adjacent pixels in a local homogeneous area have the spatial distribution consistency of land objects. To refine the spatial information, the LBP model is applied in each band for extracting texture features. The overview architecture of the proposed RAGE is shown in Fig. 1.

Suppose that the neighborhood space $\Omega(x_i)$ are expressed as $\{x_{i1}, x_{i2}, ..., x_{i(w^2-1)}\}$, where $w^2 - 1$ is the number of neighbors of center x_i. The local spatial feature of x_i by LBP extractor can be defined as:

$$\operatorname{LBP}_{w^2-1, r}(x_i) = \sum_{h=1}^{w^2-1} L(x_{ih} - x_i) 2^h, x_{ih} \in \Omega(x_i), \quad (2)$$

where r is the radius of a circle centered at x_i. $L(x_{ih} - x_i) = 1$ if $x_{ih} > x_i$, otherwise $L(x_{ih} - x_i) = 0$. As shown in Eq.(2), each neighbor is assigned as a binary label after LBP. It is a non-parametric and simple to calculate. The output is a binary code, which reflects the texture orientation and smoothness within a certain window w. Based on the LBP code, the corresponding histogram is obtained over the local patch centered at x_i. Then all spectral bands of LBP histograms are concatenated to build the spatial feature. LBP is insensitive to the monotonic illumination changes, which is suitable for single band processing in HSI.

In spectral feature extraction, we use linear prediction error (LPE) to select spectral bands with distinctive features, which is similar to principal component analysis based on band similarity. Finally, we obtain the spatial-spectral features by stacking the above LBP feature and spectral feature into one-dimensional vector.

2.3. Anchor-based Adjacent Graph Construction

After obtaining the spatial-spectral features, we randomly choose k_{ss} features to build an anchor graph in this section. To reduce the computational complexity, anchor-based strategy is adopted to learn the adjacent graph. Ideally, each sample can be represented by the linear combination of the anchors. Furthermore, the data labels are also the specific representations of the samples. Therefore, the label prediction function is denoted as

$$f(x_i) = \sum_{i=1}^{m} P_{ij} f(u_j), \quad (3)$$

where m is the number of anchors ($m \ll n$). $\mathbf{U} = [u_1, u_2, ...u_m]^T \in \mathbb{R}^{m \times d}$ denotes the anchor set. $\mathbf{P} \in \mathbb{R}^{n \times m}$ is the adjacent graph, where p_{ij} denotes the similarity between x_i and u_j. Assume that $\mathbf{F} = [f(x_1), f(x_2), ..., f(x_n)]^T$

$\in \mathbb{R}^{n \times c}$ and $\mathbf{F}_u = [f(u_1), f(u_2), ..., f(u_m)]^T \in \mathbb{R}^{m \times c}$, the Eq. (3) can be rewritten as $\mathbf{F} = \mathbf{PF}_u$. Therefore, the design of $\mathbf{P}$ is a key problem, which mainly consists of two steps:

Anchors Generation: The anchors are usually generated by k-means or random sampling. To yield more representative anchors, we employ k-means method.

Adjacent Graph Learning: The matrix $\mathbf{P}$ is constructed by k-nearest neighbors method. Inspired by [4, 12], we adopt an adaptive neighbors assignment strategy. The nearest anchors assignment of x_i is seen as the objective function:

$$\min_{p_i^T \mathbf{1}=1, 0 \leq p_{ij} \leq 1} \sum_{j=1}^{m} \|x_i - u_j\|_2^2 p_{ij} + \gamma \|p_{ij}\|_2^2, \quad (4)$$

where p_i^T is the i-th row of $\mathbf{P}$. In Eq.(4), the first part is the smoothness term, which ensures the nearby pixels belong to similar semantic labels. The second is the regularization term, which is to prevent the trivial solution of Eq.(4), where $\gamma > 0$ is a regularization parameter. Let us define $e_{ij} = \|x_i - u_j\|_2^2$, then the objective function (4) is converted into

$$\min_{p_i^T \mathbf{1}=1, 0 \leq p_{ij} \leq 1} \frac{1}{2} \left\| p_i + \frac{1}{2\gamma} e_i \right\|_2^2. \quad (5)$$

According to the convex optimization of the Lagrangian function, we have $\gamma = \frac{k}{2} e_{i,k+1} - \frac{1}{2} \sum_{j=1}^{k} e_{ij}$, where k is the number of nonzero values. By this way, we achieve the optimal solution p_{ij}^* as follows:

$$p_{ij}^* = \frac{e_{i,k+1} - e_{ij}}{k e_{i,k+1} - \sum_{j'=1}^{k} e_{ij'}}. \quad (6)$$

After getting the adjacent graph $\mathbf{P}$, the normalized adjacent graph $\mathbf{S}$ can be computed as $\mathbf{S} = \mathbf{P} \boldsymbol{\Lambda}^{-1} \mathbf{P}^T$, where $\boldsymbol{\Lambda}$ is a diagonal matrix whose j-th element is represented as $\Lambda_{jj} = \sum_{i=1}^{n} p_{ij}$, and $\boldsymbol{\Lambda} \in \mathbb{R}^{m \times m}$. Intuitively, the element s_{ij} of matrix $\mathbf{S}$ is expressed as $s_{ij} = p_i^T \Lambda^{-1} p_j$ that satisfies $s_{ij} = s_{ji}$. Moreover, it is easy to prove that matrix $\mathbf{S}$ is positive semidefinite and doubly stochastic. The above property of $\mathbf{S}$ is crucial for optimizing AG-based SSL model, and the details will be given later.

2.4. AG-based SSL Model

The graph-based SSL objective function has described in Eq. (1). As mentioned above, the prediction function of labels is represented as $\mathbf{F} = \mathbf{PF}_u$. To infer the labels of unlabeled samples in HSI classification, the AG-based SSL model based on section 2.3 can be considered as the following problem:

$$\mathcal{L}(\mathbf{F}_u) = \text{Tr}\big((\mathbf{F}_l - \mathbf{Y}_l)^T (\mathbf{F}_l - \mathbf{Y}_l)\big) + \alpha \text{Tr}(\mathbf{F}^T \mathbf{LF}), \quad (7)$$

where $\mathbf{F}_l = \mathbf{P}_l \mathbf{F}_u$, the $\mathbf{F}_l$ and $\mathbf{P}_l$ are the sub-matrix of $\mathbf{F}$ and $\mathbf{P}$, respectively. $\mathbf{Y}_l$ is the labels of labeled samples, $\| \cdot \|_F$

is the Frobenius norm. The α is the regularization parameter. The Eq. (7) can be represented as

$$\begin{aligned} \mathcal{L}(\mathbf{F}_u) &= \|\mathbf{P}_l \mathbf{F}_u - \mathbf{Y}_l\|_F^2 + \alpha \text{Tr}((\mathbf{PF}_u)^T \mathbf{L}(\mathbf{PF_u})) \\ &= \|\mathbf{P}_l \mathbf{F}_u - \mathbf{Y}_l\|_F^2 + \alpha \text{Tr}(\mathbf{F}_u{}^T \mathbf{L}_A \mathbf{F}_u), \end{aligned} \quad (8)$$

where $\mathbf{L}_A = \mathbf{P}^T \mathbf{LP}$, and $\mathbf{L} = \mathbf{D} - \mathbf{S}$. According to the aforementioned property of $\mathbf{S}$, we have

$$d_{ii} = \sum_{j=1}^{n} s_{ij} = \sum_{j=1}^{n} p_i^T \Lambda^{-1} p_j = p_i^T \sum_{j=1}^{n} \Lambda^{-1} p_j = p_i^T \mathbf{1} = 1. \quad (9)$$

Therefore, according to $\mathbf{L} = \mathbf{D} - \mathbf{S} = \mathbf{I} - \mathbf{S} = \mathbf{I} - \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}^T$, we can obtain

$$\mathbf{L}_A = \mathbf{P}^T(\mathbf{I} - \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P^T})\mathbf{P} = \mathbf{P^T P} - (\mathbf{P^T P})\boldsymbol{\Lambda}^{-1}(\mathbf{P^T P}). \quad (10)$$

The optimization problem (8) can be solved by setting the first derivative as zero, and the final solution is

$$\mathbf{F}_u^* = (\mathbf{P}_l^T \mathbf{P}_l + \alpha \mathbf{L}_A)^{-1} \mathbf{P}_l^T \mathbf{Y}_l. \quad (11)$$

After obtaining the $\mathbf{F}_u^*$, the predicted label of unlabeled pixel x_i can be determined by

$$y_i = \arg \max_{j \in \{1, ..., c\}} \mathbf{P}_i . \mathbf{F}_{u_j}^*. \quad (12)$$

Actually, we will build k_g graphs by repeating section 2.3. Based on ensemble learning strategy, the labels of unlabeled samples are determined by the voting results of k_g graphs acquired from section 2.4.

3. EXPERIMENTS

3.1. Datasets and Experimental Setup

To demonstrate the effectiveness of the proposed RAGE method, the experiments are conducted on public Indian Pines and Pavia University datasets. Indian Pines dataset contains 16 land cover types, which leaves 200 bands to be used for experiments after removing noise and water absorption bands. Pavia University includes 9 classes and 103 bands for classification task. The proposed RAGE is compared with R-VCANet [3], LSLRR [2], SS-RMG [13] and SSHGDA [10]. The parameters of RAGE are $\{\alpha = \alpha_l = 0.01, \alpha_v = 10^{-6}, w = 7, k_{ss} = 96, k_g = 4\}$. We randomly choose 5% and 1% of Indian Pines and Pavia University datasets for training, respectively. Three quantitative metrics are used to evaluate performance, including overall accuracy (OA), average accuracy (AA) and Kappa coefficient.

3.2. Experimental Results and Analyses

The visual and quantitative results of the methods are given in Fig. 2, Table 1 and Table 2. It can be revealed that RAGE obtains better classification results and the higher accuracy than competitors from the perspective of OA, AA and Kappa.
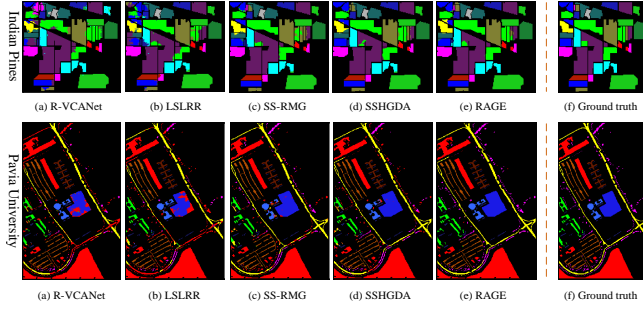
(a) R-VCANet (b) LSLRR (c) SS-RMG (d) SSHGDA (e) RAGE (f) Ground truth

(a) R-VCANet (b) LSLRR (c) SS-RMG (d) SSHGDA (e) RAGE (f) Ground truth

Fig. 2. The classification maps of different methods.

By adopting LBP feature extractor, the distinctive spatial features are obtained as shown in Fig. 3, which discovery the fine differences from different bands and enrich the local structure information. This enhance the discriminability of learners and effectively improve the classification performance of the proposed RAGE. Moreover, we randomly choose several subsets of features to build different anchor graphs in parallel, then obtaining the best prediction from multiple learners by voting. By this way, the classification accuracy and efficiency of the RAGE are further improved with small training samples. From Table 1 and Table 2, RAGE consumes the least running time on two datasets, and its accuracy is still maintained at a good level comparing with other algorithms.

Table 1. Quantitative metrics of different methods on Indian Pines dataset. The optimal value is highlighted in bold.

Methods	OA(%)	AA(%)	Kappa	Time(s)
R-VCANet	88.09	90.97	0.8639	1378.5
LSLRR	90.97	92.65	0.8966	336.2
SS-RMG	96.89	**97.73**	0.9646	63.9
SSHGDA	97.78	96.22	0.9747	1029.7
RAGE	**98.75**	96.90	**0.9857**	**14.2**

Table 2. The results of methods on Pavia University dataset.

Methods	OA(%)	AA(%)	Kappa	Time(s)
R-VCANet	85.78	82.31	0.8002	2789.7
LSLRR	91.27	86.89	0.8809	1257.3
SS-RMG	92.71	87.73	0.9009	380.9
SSHGDA	93.53	89.29	0.9123	13938.1
RAGE	**99.26**	**98.98**	**0.9902**	**179.5**

4. CONCLUSION

In this paper, we propose an efficient spatial-spectral HSI classification method based on multiple random anchor graphs ensemble learning. By adopting LBP model, the fine spatial features are obtained, such that the discriminability is enhanced. The multi-graphs with random spatial-spectral features are built in parallel, which further improves the efficiency of the proposed model and learns the diversity of HSI data. The adaptive neighbors assignment in AG construction decreases the computational complexity. Extensive experiments on two public HSI datasets verify the effectiveness and advantages of the proposed RAGE.

5. ACKNOWLEDGEMENT

(a) Band 2 (b) Band 30 (c) Band 60 (d) Band 100
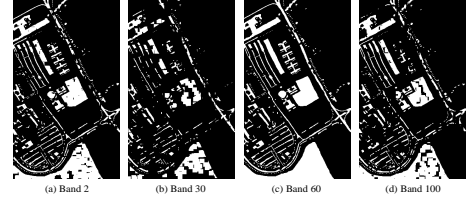
Fig. 3. The spatial features based on LBP on bands.

6. REFERENCES

[1] P. Sellars, A. I. Aviles-Rivero, and C. Schnlieb, "Superpixel contracted graph-based learning for hyperspectral image classification," *IEEE TGRS*, vol. 58, no. 6, pp. 4180–4193, 2020.

[2] Q. Wang, X. He, and X. Li, "Locality and structure regularized low rank representation for hyperspectral image classification," *IEEE TGRS*, vol. 57, no. 2, pp. 911–923, 2019.

[3] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *IEEE IGARSS*, 2015, pp. 4959–4962.

[4] F. Nie, S. Shi, and X. Li, "Semi-supervised learning with auto-weighting feature and adaptive graph," *IEEE TKDE*, vol. 32, no. 6, pp. 1167–1178, 2019.

[5] A. Anis, A. El Gamal, and A. Avestimehrand A. Ortega, "A sampling theory perspective of graph-based semi-supervised learning," *IEEE TIT*, vol. 65, no. 4, pp. 2322–2342, 2018.

[6] F. He, R. Wang, and W. Jia, "Fast semi-supervised learning with anchor graph for large hyperspectral images," *Pattern Recognit. Lett.*, vol. 130, pp. 319–326, 2020.

[7] H. Zhang, H. Zhai, L. Zhang, and P. Li, "Spectral-spatial sparse subspace clustering for hyperspectral remote sensing images," *IEEE TGRS*, vol. 54, no. 6, pp. 3672–3684, 2016.

[8] H. Huang, G. Shi, H. He, Y. Duan, and F. Luo, "Dimensionality reduction of hyperspectral imagery based on spatial–spectral manifold learning," *IEEE TCYB*, vol. 50, no. 6, pp. 2604–2616, 2019.

[9] H. Huang, Y. Duan, H. He, and G. Shi, "Local linear spatial-spectral probabilistic distribution for hyperspectral image classification," *IEEE TGRS*, vol. 58, no. 2, pp. 1259–1272, 2020.

[10] F. Luo, B. Du, L. Zhang, L. Zhang, and D Tao, "Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image," *IEEE TCYB*, vol. 49, no. 7, pp. 2406–2419, 2018.

[11] Y. Zhang, G. Cao, A. Shafique, and P. Fu, "Label propagation ensemble for hyperspectral image classification," *IEEE JSTARS*, vol. 12, no. 9, pp. 3623–3636, 2019.

[12] X. Li, M. Chen, F. Nie, and Q. Wang, "A multiview-based parameter free framework for group detection," in *AAAI*, 2017, pp. 4147–4153.

[13] F. Gao, Q. Wang, Y. Dong, and Q. Xu, "Spectral and spatial classification of hyperspectral images based on random multi-graphs," *Remote Sens.*, vol. 10, no. 8, pp. 1271, 2018.