2024 14th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS) | 979-8-3315-1313-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/WHISPERS65427.2024.10876482

LEARNING BY EVOLVING NONLINEAR DIFFUSION FOR ACTIVE LEARNING ON HYPERSPECTRAL IMAGES

Kabir Tripathi, James M. Murphy
Tufts University, Department of Mathematics
Medford, MA 02155, USA

ABSTRACT

A method for active learning of hyperspectral images (HSI) is proposed, which leverages time-evolving diffusion processes on graphs to determine query points and propagate their labels to the full data set. Initially, a diffusion process is defined on a graph in which each HSI pixel is a node, with edge connection strength scaling with the distance between pixels in the spectral domain. At each stage of the iterative sampling process, queried labels are used to update the underlying diffusion matrix by weakening edges between points in different classes, thereby infusing the intrinsic representation encoded in the diffusion matrix with revealed label information. The proposed method, *Learning by Evolving Nonlinear Diffusion (LEND)*, combines robust performance with the mathematical tractability of diffusion geometry, leading to superior labeling accuracy with fewer labeled samples compared to a baseline in which the underlying matrix is not iteratively updated. Experiments on the Salinas A HSI demonstrate the effectiveness and efficiency of LEND.

***Index Terms*—** active learning, semisupervised learning, hyperspectral imaging, diffusion geometry

1. INTRODUCTION

Machine learning has provided revolutionary new tools for remote sensing [1, 2, 3], but state-of-the-art methods often require huge labeled training sets. In particular, supervised deep learning methods can achieve near-perfect labeling accuracy on high-dimensional hyperspectral images (HSI), provided large collections of labeled pixels are available [4]. However, the practicality of these methods may be limited in settings when data is collected at a pace that exceeds human ability to generate corresponding labeled training data.

In order to account for this, methods that require only a very small number of labels are needed [5]. Typically, in a supervised scheme, labeled points are split into training and testing sets. However, this approach is often impractical, as obtaining sufficient labels for both training and testing can be costly, especially in hyperspectral imaging (HSI). The *active learning* regime is particularly attractive for HSI labeling problems [6, 7, 5, 8, 9, 10, 11, 12]. In active learning, an algorithm is provided with an unlabeled dataset, and the algorithm iteratively queries points for labels. The goal of query selection is to maximize the efficiency of labeling under a given budget. Two primary strategies include hypothesis space reduction, which aims to quickly converge to an optimal classifier by choosing influential points, and cluster exploitation, which focuses on sampling in complex regions while avoiding redundant labeling in homogeneous areas [9]. Our proposed method aligns with the second strategy, utilizing the geometry of data to identify key points for labeling. By leveraging the diffusion geometry of the data, our method robustly handles high-dimensional, non-linear, and noisy data. This enhances the selection of impactful query points, improving classification accuracy with fewer labeled samples.

We propose a new active learning method for HSI pixel labeling called *Learning by Evolving Nonlinear Diffusion (LEND)*. LEND iteratively refines labels using the Learning by Active Nonlinear Diffusion (LAND) algorithm [13, 6]. LEND updates the diffusion matrix by weakening edges between different classes (as per labels given by LAND), and recomputes a diffusion-based embedding at each update. This process acts as a method of denoising the diffusion eigenmaps, achieving superior classification accuracy with fewer labeled samples. By learning an accurate representation of the data, LEND is better informed on how to query for labels. Experiments on the Salinas A HSI demonstrate the effectiveness and efficiency of LEND and show significant improvements in labeling accuracy over the existing active learning method LAND [6].

The rest of the paper is organized as follows. In Section 2, we review related works in active learning for HSI and discuss diffusion geometry and LAND [6]. Following that we outline LEND in Section 3 and provide empirical results in Section 4.

2. BACKGROUND

Before delving into our proposed method, we will review similar approaches to active learning for HSI and discuss prerequisite details regarding diffusion geometry and LAND.

Overview of Active Learning for HSI: Active learning is a learning paradigm where the user selects the training data [5]. The underlying idea is that a few informative training sam-

ples can be sufficient for training an algorithm and achieving accurate results. This framework has been applied in remote sensing for HSI image classification [6, 14, 15, 16, 17, 18]. A group of methods known as deep clustering aim to learn data features and perform clustering simultaneously, demonstrating strong empirical results [19, 20, 21]. For HSI images, multiple studies have used different deep learning architectures to extract essential features for downstream tasks such as classification [16, 17, 22, 18, 23, 15]. In contrast to deep learning methods, our algorithm's embedding scheme enjoys interpretability and efficient computational complexity. Recent works have used a variational autoencoder to extract features from the data, and then use diffusion eigenmaps to extract features once again, and then run an inference procedure to classify the data [15].

The main idea in this paper is that the active learning process depends on the representation and geometry of the data as highlighted in LAND and its unsupervised analogue Learning by Unsupervised Nonlinear Diffusion (LUND) [14]. We seek to learn a low-dimensional, nonlinear manifold the data lies on. By approximating a metric on this manifold, we can calculate certain test statistics (detailed in [6]) to identify modes in the data. After calculating said modes, we iteratively label the data using the manifold's metric.

Background on Diffusion Geometry and LAND: We represent hyperspectral images as $X = \{x_i\}_{i=1}^n \subset \mathbb{R}^N$ where the pixels are points in $\mathbb{R}^N$ and N is the number of spectral bands. Define $NN_k(x_i)$ as the set of k-nearest neighbors of x_i in X under the ℓ_2 metric. Let $W \in \mathbb{R}^{nxn}$ be the weight matrix with $W_{ij} = \exp(-\|x_i - x_j\|_2^2/\sigma^2), x_j \in NN_k(x_i)$ with σ denoting a scale parameter. With this, the notion of the degree of x_i naturally follows as $\deg(x_i) := \sum_{x_j \in X} W_{ij}$. To define a random walk on X, normalize W to form the $n \times n$ transition matrix P with $P_{ij} = W_{ij}/\deg(x_i)$ [24]. It is shown in [24] that P admits a spectral decomposition $\{(\lambda_\ell, \Psi_\ell)\}_{\ell=1}^n$ where without loss of generality $1 = \lambda_1 \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. Let π be the stationary distribution of P, i.e. $\pi = \pi P$. The *diffusion distance at time* t between $x_i, x_j \in X$ is defined as $D_t(x_i, x_j) = \sqrt{\sum_{k=1}^n (p_t(x_i, x_k) - p_t(x_j, x_k))^2 \frac{1}{\pi(x_k)}} = \sqrt{\sum_{\ell=1}^n \lambda_\ell^{2t} (\Psi_\ell(x_i) - \Psi_\ell(x_j))^2}$ [24]. Note that t tells us how long the diffusion process runs. In this paper, we use $t = 30$ for experiments as it allows for the process to run long enough that it reveals structure in the data without destroying it. Diffusion distances measure how far away are two points by capturing how different their respective transition probabilities are to other points. So if their random walks are similar, the two points are close to each other in diffusion space.

A fundamental aspect of our active learning algorithm is to use diffusion distances on X to identify points to query for labels. LAND uses a kernel density estimator (KDE) and diffusion geometry [24, 25] for this task. The KDE is defined as $p(x) = \sum_{y \in NN_k(x)} \exp(-\|x - y\|_2^2/\sigma_0^2)$ with σ_0 denoting a scale parameter. We use 500 nearest neighbors and σ_0 equals the mean of the nonzero pairwise distances across the dataset. For $x \in X$, let

$$\rho_t(x) = \begin{cases} \min\limits_{p(y) \geq p(x), x \neq y} D_t(x, y), & x \neq \operatorname{argmax}\limits_z p(z), \\ \max\limits_{y \in X} D_t(x, y), & x = \operatorname{argmax}\limits_z p(z), \end{cases}$$

be the nearest neighbor in diffusion space of higher density [6]. The maximizers of $\mathcal{D}_t(x) = p(x)\rho_t(x)$ are queried for labels [6]. The maximizers are in high density regions and far from other points in diffusion space, so they can be thought of as modes of the data set. These labels are then propagated to other data points by proceeding from high to low density and assigning each unlabeled point the same label as its D_t-nearest neighbor of higher density that is labeled; see [6] for a detailed explanation. An intuitive way to visualize the algorithm is to imagine the modes (see Figure 1) diffusing out into low density regions labeling points that haven't been labeled yet.

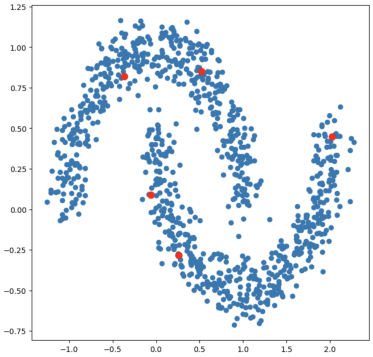

Fig. 1: Query points shown in red on the nonlinear two moons dataset.

3. PROPOSED ALGORITHM

We propose an active learning algorithm, LEND, that uses a density-based diffusion embedded scheme to improve the diffusion map used to estimate the structure of the data, ultimately improving our low-dimensional estimation of the data and the points we query for labels (as determined by $\mathcal{D}_t(\cdot)$).

Our model operates through an alternating algorithm, switching between estimating the underlying diffusion map and classifying the data. Initially, we utilize the LAND spectral classification scheme [6]—described in Algorithm 1 —to obtain preliminary labels. Using these labels, we strategically weaken edges in the diffusion map and restart the labeling process.

In both LEND and LAND, we construct the diffusion graph using a sparse weight matrix W. We achieve this by

selecting the 100 nearest neighbors of each point based on the ℓ_2 (Euclidean) metric and assigning weights to the edges using the Gaussian kernel. This step is crucial for effective edge weakening, as it allows the removal of a single edge to potentially disconnect two clusters.

LEND—detailed in Algorithm 2—operates through an alternating algorithm that switches between estimating the underlying diffusion map and classifying the data. Initially, we utilize the LAND spectral classification scheme to obtain preliminary labels. Using these labels, we weaken edges in the diffusion map between points of different estimated labels and restart the labeling process with the new diffusion map.

Two key parameters in our LEND algorithm are J and α. The parameter J determines the number of classification-diffusion estimation iterations. We start our first iteration of LAND at the total budget minus J to avoid exhausting all our queries at the beginning. This approach allows us to continuously query from the start to the end of the list of $\mathcal{D}_t(\cdot)$ maximizers or limit our queries to the points budgeted for each iteration.

In our implementation, we query all points budgeted for a given iteration and use those for inference. Importantly, we experimentally found that this approach does not exceed the allocated budget. If one aims to query K points, it can sometimes be beneficial to provide the algorithm with slightly fewer than K points initially. By allowing the algorithm to freely query all budgeted points at each iteration, it ultimately queries a total of K points. This approach operates under the assumption that early maximizers of $\mathcal{D}_t(\cdot)$ remain consistent. An alternative procedure could involve continuously querying points until the budget is exhausted.

The parameter α controls the scaling of edges between points in different classes. A higher α provides resilience against early misclassifications, allowing the algorithm to make use of the information from LAND's predictions, even if they are initially inaccurate. This balance helps in leveraging the evolving information to refine the diffusion map progressively, enhancing the overall classification accuracy through the iterative process.

4. EXPERIMENTAL RESULTS

We demonstrate the accuracy of LEND experimentally on the Salinas A HSI [26]. We implement both LAND and LEND in Python using Numpy and Scipy [1]. The (full) Salinas scene was captured over Salinas Valley, California. The image has a spatial resolution of 3.7-meter pixels and contains 224 spectral bands. The ground truth has 16 classes. We consider the Salinas A dataset, which is a 6-class subset of the Salinas dataset. Figure 2 shows a low-dimensional visualization of the data and the ground truth labels. The performance of the algorithm is assessed using overall accuracy, defined as the ratio of correctly estimated labels to total number of labels after aligning via the Hungarian algorithm.

[1] https://github.com/kabirst11/LEND

Algorithm 1: Learning by Active Nonlinear Diffusion (LAND)

Input: $\{x_i\}_{i=1}^n$ (Unlabeled Data); $\{p(x_i)\}_{i=1}^n$ (Kernel Density Estimate); P (Diffusion Matrix); t (Time Parameter); B (Budget); $\mathcal{O}$ (Labeling Oracle);
Output: Y (Labels)

Compute $\{(\lambda_\ell, \Psi_\ell)\}_{\ell=1}^M$, the spectral decomposition of P.
Compute $\{\rho_t(x_i)\}_{i=1}^n$.
Compute $\mathcal{D}_t(x_i) = p(x_i)\rho_t(x_i)$.
Sort the data in order of decreasing $\mathcal{D}_t$ value to acquire the ordering $\{x_{m_i}\}_{i=1}^n$.
for $i = 1 : B$ **do**
    Query $\mathcal{O}$ for the label $L(x_{m_i})$ of x_{m_i}.
    Set $Y(x_{m_i}) = L(x_{m_i})$.
end for
Sort X according to $p(x)$ in decreasing order as $\{x_{\ell_i}\}_{i=1}^n$.
for $i = 1 : n$ **do**
    if $Y(x_{\ell_i}) = 0$ **then**
        $Y(x_{\ell_i}) = Y(z_i)$, $z_i = \underset{z}{\operatorname{argmin}}\{D_t(z, x_{\ell_i}) \mid p(z) > p(x_{\ell_i})$ and $Y(z) > 0\}$.
    end if
end for

Algorithm 2: Learning by Evolving Nonlinear Diffusion (LEND)

Input: $\{x_i\}_{i=1}^n$ (Unlabeled Data); $\{p(x_i)\}_{i=1}^n$ (Kernel Density Estimate); P (Diffusion Matrix); t (Time Parameter); B (Budget); $\mathcal{O}$ (Labeling Oracle); J (Number of Iterations); α (Scale at which to weaken edges)
Output: Y_J (Labels)

Initialize diffusion matrix $P_1 = P$.
Initialize matrix $\mathcal{B}$
for $j = 1 : J$ **do**
    $Y_j = \text{LAND}(P_j, \text{Budget} = B - (J - j))$ *(Other parameters are default, hidden for space)*
    Evaluate $\mathcal{B}$ according to Y_j, $\mathcal{B}_{ik} = 1$ if $Y_j(x_i) = Y_j(x_k)$ and α otherwise
    $\hat{P}_{j+1} = \mathcal{B} \odot P_j$ (Entry wise multiplication, to weaken all edges between all points in different classes)
    Normalize rows of $\hat{P}_{j+1}$ so they sum to 1, denote this as P_{j+1}.
end for

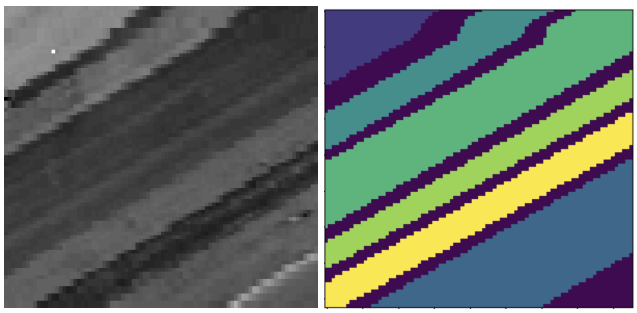

Fig. 2: The 83×86 Salinas A HSI data consists of 6 classes. *Left:* the sum of all spectral bands. *Right:* the ground truth.

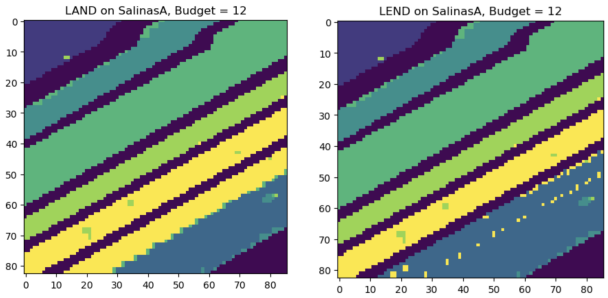

Fig. 4: Comparing the predictions given by LAND and LEND on Salinas A, with a budget of 12 labels.

The Salinas A HSI dataset, sized 83 x 86 x 224, is represented as a point cloud of 7138 x 224. In LEND we exhibit 98% accuracy with a budget of 12, almost an 11% improvement on LAND with the same budget and with other parameters set the same; see Figures 3 and 4 for comprehensive results. We see moreover very fast convergence to near-perfect accuracy with LEND, compared to LAND. For our experiments we use $J = 3$ and $\alpha = 0.75$. Both J, α were chosen by a binary search. This choice of J gave the algorithm enough labels to start with so that the inference procedure is informative while leaving enough room in the budget to pick better query points in subsequent iterations. This choice of α is larger because it ensures weak accuracy early on doesn't cause premature, complete edge deletion.

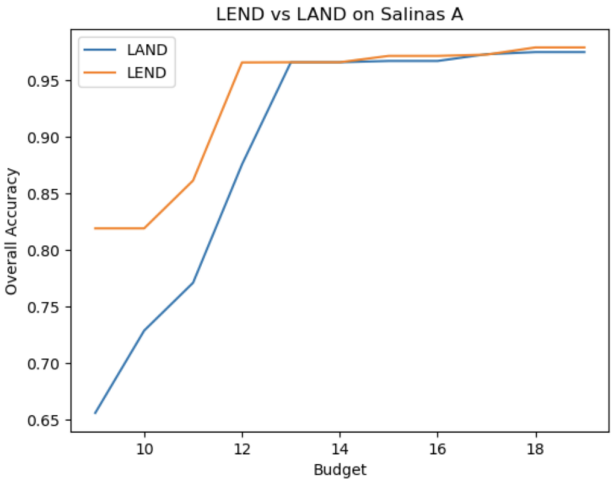

Fig. 3: Comparing the performance of LAND and LEND on Salinas A.

Complexity and run time: LAND has time complexity $O(C_{NN} + nK_{NN} + nlog(n))$ where C_{NN} is the cost of computing all K_{NN} nearest neighbors [6]. LEND is essentially LAND plus calculating the matrix $\mathcal{B}$ after each iteration, which has $O(n^2)$ operations, so the complexity of LEND is $O(C_{NN} + nK_{NN} + nlog(n) + n^2)$. All simulations were run on a 2019 MacBook Pro, with 8 GB of memory and a Quad-Core Intel i5 processor. LAND takes about 49 seconds to run and LEND takes about 2 minutes 51 seconds. LEND has to perform a spectral decomposition J times, every time P is updated and put into LAND, for this reason it is slower than LAND, scaling at a rate linear in J.

5. CONCLUSIONS AND FUTURE DIRECTIONS

The proposed active learning algorithm, LEND, provides a substantial improvement over LAND in the low query regime. We use LAND's inference procedure to self-correct the diffusion matrix, thereby giving us a better representation of the data and improving which points we query for labels. In future work, we seek to create a more principled way of enforcing budget constraints in Algorithm 2. We also seek to study how inference can inform adding edges to the graph and create a data model where these inference-informed diffusion constructions are most useful. Moreover, understanding the impact of α is a topic of interest, and we expect larger α to be more appropriate in the case the preliminary LAND labeling is poor.

Acknowledgments: The authors acknowledge partial support from NSF DMS 2309519, NSF DMS 2318894, and the Kokulis Family Endowed Summer Scholars Fund.

6. REFERENCES

[1] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, "A survey: Deep learning for hyperspectral image classification with few labeled samples," *Neurocomputing*, vol. 448, pp. 179–204, 2021.

[2] H. Zhai, H. Zhang, P. Li, and L. Zhang, "Hyperspectral image clustering: Current achievements and future lines," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 4, pp. 35–67, 2021.

[3] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. Benediktsson, "Deep learning for hyperspectral im-

age classification: An overview," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6690–6709, 2019.

[4] X.X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, 2017.

[5] D.A. Cohn, Z. Ghahramani, and M.I. Jordan, "Active learning with statistical models," in *Advances in Neural Information Processing Systems*, 1995.

[6] M. Maggioni and J.M. Murphy, "Learning by active nonlinear diffusion," *Foundations of Data Science*, vol. 1, no. 3, 2019.

[7] J.M. Murphy and M. Maggioni, "Iterative active learning with diffusion geometry for hyperspectral images," in *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2018, pp. 1–5.

[8] S. Dasgupta, D.J. Hsu, and C. Monteleoni, "A general agnostic active learning algorithm," in *Advances in Neural Information Processing Systems*, 2008, pp. 353–360.

[9] S. Dasgupta, "Two faces of active learning," *Theoretical Computer Science*, vol. 412, no. 19, pp. 1767–1781, 2011.

[10] M.-F. Balcan, A. Broder, and T. Zhang, "Margin based active learning," in *International Conference on Computational Learning Theory*. 2007, pp. 35–50, Springer.

[11] M.-F. Balcan, A. Beygelzimer, and J. Langford, "Agnostic active learning," *Journal of Computer and System Sciences*, vol. 75, no. 1, pp. 78–89, 2009.

[12] R.M. Castro and R.D. Nowak, "Minimax bounds for active learning," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2339–2353, 2008.

[13] J.M. Murphy and M. Maggioni, "Unsupervised clustering and active learning of hyperspectral images with nonlinear diffusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, 2019.

[14] M. Maggioni and J.M. Murphy, "Learning by unsupervised nonlinear diffusion," *Journal of Machine Learning Research*, vol. 20, no. 160, 2019.

[15] Abiy Tasissa, Duc Nguyen, and James M. Murphy, "Deep diffusion processes for active learning of hyperspectral images," in *IEEE International Geoscience and Remote Sensing Symposium*, 2021, pp. 3665–3668.

[16] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, 2014.

[17] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, 2016.

[18] M.E. Paoletti, J..M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, 2019.

[19] F. Tian, B. Gao, Q. Cui, E. Chen, and T.-Y. Liu, "Learning deep representations for graph clustering," in *AAAI Conference on Artificial Intelligence*, 2014.

[20] C. Song, F. Liu, Y. Huang, L. Wang, and T. Tan, "Auto-encoder based data clustering," in *Iberoamerican Congress Pattern Recognit.* Springer, 2013.

[21] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in *International Conference on Machine Learning*, 2016.

[22] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sensing*, vol. 9, no. 1, 2017.

[23] M. He, B. Li, and H. Chen, "Multi-scale 3D deep convolutional neural network for hyperspectral image classification," in *International Conference on Image Processing*. IEEE, 2017.

[24] R.R. Coifman and S. Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, 2006.

[25] R.R. Coifman, S. Lafon, A.B. Lee, M. Maggioni, B. Nadler, F. Warner, and S.W. Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps," *Proceedings of the National Academy of Sciences*, vol. 102, no. 21, 2005.

[26] "Hyperspectral remote sensing scenes," https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes, Accessed: 2024-4-11.