Fractal Autoencoder Based supervised Hyperspectral Bands Selection for Remote Sensing Land-Cover Classification.†

Sara BENALI[1], Mohammed El Amine LARABI[2] and Dalila ATTAF[3].

Ecole Nationale Supérieure des Sciences Geodésique et Techniques Spatiales1; benalisara931@gmail.com

Algerian Spatial Agency2; arabi.mohamed.lamine@gmail.com

Algerian Spatial Agency3; dalila.attaf@gmail.com

Abstract

Band selection is a frequently used dimension reduction technique for hyperspectral images (HSI) to address the "curse of dimensionality" phenomenon in machine learning (ML). This technique identifies and selects a subset of the most important bands from the original ones to remove redundancy and noisy information while maintaining optimal generalization ability. Band selection methods can be categorized into supervised and unsupervised techniques depending on whether labels are used. Unsupervised band selection and feature extraction framework is proposed in this study. The framework trains a sub-neural network to identify the most important and informative bands from the original data space, which is then projected to a reduced and more informative feature space. The classification performance of the selected bands combination on the Pavia University HSI datasets has been verified using multiple machine learning algorithms. The proposed method not only enhances the classification results of HSI, but also reduces the computational time and data storage requirements compared to other state-of-the-art band selection approaches.

Keywords Keywords: Hyperspectral Images,Supervised Band Selection, Deep Learning, Autoencoders, Classification.

1. Introduction

High-dimensional datasets are common in various fields, such as image processing, genomics, finance, and more. These datasets have a wide range of features (attributes), often surpassing the number of samples available for analysis. This wealth of information is valuable, but it also presents numerous challenges, collectively known as the "curse of dimensionality". The later involves issues such as increased computational complexity,overfitting, degraded model performance and reduced interpretability. These challenges hinder the effectiveness of traditional data analysis methods[1]. Feature selection is a technique that identifies a subset of relevant features from a high-dimensional dataset. There are three main types of feature selection methods: filter methods, wrapper methods, and embedded methods.

Filter methods assess feature importance independently of any specific learning algorithm, while wrapper methods use a specific learning algorithm to evaluate the impact of feature subsets on model performance[2]. Embedded methods combine feature selection

[1] Bandos, Bruzzone, et Camps-Valls, « Classification of Hyperspectral Images With Regularized Linear Discriminant Analysis ».

[2] Wah et al., « Feature Selection Methods: Case of Filter and Wrapper ».

seamlessly with the learning process itself. The choice of feature selection method depends on the specific application and the available resources.

These methods cover a wide range of feature selection techniques, including Principal Feature Analysis (PFA)[3], which prioritizes key features through statistical measures; Multi-Cluster Feature Selection (MCFS)[4], which leverages clustering techniques; Unsupervised Discriminative Feature Selection (UDFS), which seeks to maximize feature discrimination; and Principal Component Analysis (PCA), which focuses on orthogonal transformations. However, a significant challenge arises when the selected features exhibit high correlations. This can potentially lead to the representation of only partial information and limit the global representativeness of the feature subset.

Hyperspectral imagery captures a wide range of electromagnetic radiation in hundreds of narrow bands, providing detailed information about the materials present in a scene. This makes it ideal for applications such as material identification, target detection, and environmental monitoring. However, hyperspectral imagery can be expensive and difficult to process. Multispectral imagery captures a smaller range of electromagnetic radiation in a few broad bands. This provides less detailed information, but it is more cost-effective and easier to process. Multispectral imagery is often used for applications such as land cover classification, vegetation mapping, and urban planning.

In hyperspectral remote sensing, feature or band selection and data compression are essential techniques for handling big data volumes, enhancing analysis efficiency, and simplifying the storage and transmission of hyperspectral data. This allows for more effective applications in fields like agriculture, mineral exploration, and environmental monitoring[5].

Deep learning-based feature selection methods, such as autoencoders, use neural networks to automatically identify and extract the most important features from complex datasets. Autoencoders are a type of neural network that can learn compact representations of input data. This makes them well-suited for both feature selection and data compression tasks. By learning compact representations of input data, autoencoders can enhance data analysis efficiency and preserve vital information across diverse domains [6].

A new framework for feature selection based on FAE is introduced in this paper. FAE seeks to achieve optimal feature subsets that effectively balance the representation of information and diversity, which can enhance the performance of subsequent data analysis tasks. In the following sections, the details of FAE are delved into, its unique characteristics are showcased, and its effectiveness is demonstrated through experiments and comparisons with state-of-the-art methods[7].

This work is organized as follows: In Section I, a detailed presentation of the architecture and formulation of Autoencoders and Fractal Autoencoders is provided. Following that, in Section II, our methodology for utilizing these techniques is elucidated, and our comparative analysis against several other methods is discussed.

2.Methodology

The approach tailored specifically for HSI analysis based on the concept of FAE is introduced in this section. The approach builds upon the foundation of AE while tailoring its structure to address the unique challenges posed by HSI. Figure (1) illustrates the architecture of FAE, which serves as the fundamental building block of the approach. The

[3] Lu et al., « Feature Selection Using Principal Feature Analysis ».

[4] Cai, Zhang, et He, « Unsupervised Feature Selection for Multi-Cluster Data ».

[5] Bioucas-Dias et al., « Hyperspectral Remote Sensing Data Analysis and Future Challenges ».

[6] Abid, Balin, et Zou, « Concrete Autoencoders: Differentiable Feature Selection and Reconstruction ».

[7] Wu et Cheng, « Fractal autoencoders for feature selection ».

architecture is carefully designed to facilitate feature selection in the context of hyperspectral imagery. In the subsequent sections, an in-depth explanation of the architecture and its individual components is provided.

2.1. Formalization of Autoencoders:

For hyperspectral data, we formalize the AE as follows:

$$\mathbf{min}\ ||\mathbf{X}\ - \mathbf{f}(\mathbf{g}(\mathbf{X}))||_{\mathbf{F}}^{\mathbf{2}} \qquad \mathbf{(1)}$$

Here, the encoder is represented by 'g', and the decoder is represented by 'f'. The function 'g(X)' transforms the input HSI data X into a latent space 'Rn×d', where 'd' signifies the dimension of the bottleneck layer within the AE. To illustrate, the application of our approach to a HSI dataset is considered. In the context of HSI analysis, this formalization allows the essential spectral information to be effectively captured and represented within a reduced-dimensional latent space.

2.2. Formalization Fractal Autoencoder :

FAE, a novel approach designed to tackle feature selection, introduces a concept akin to self-similarity in its operation. The primary objective of FAE is to select a subset of 'k' informative features from a hyperspectral dataset 'X', such that the chosen features collectively retain as much information about the overall spectral content of the original samples as possible.

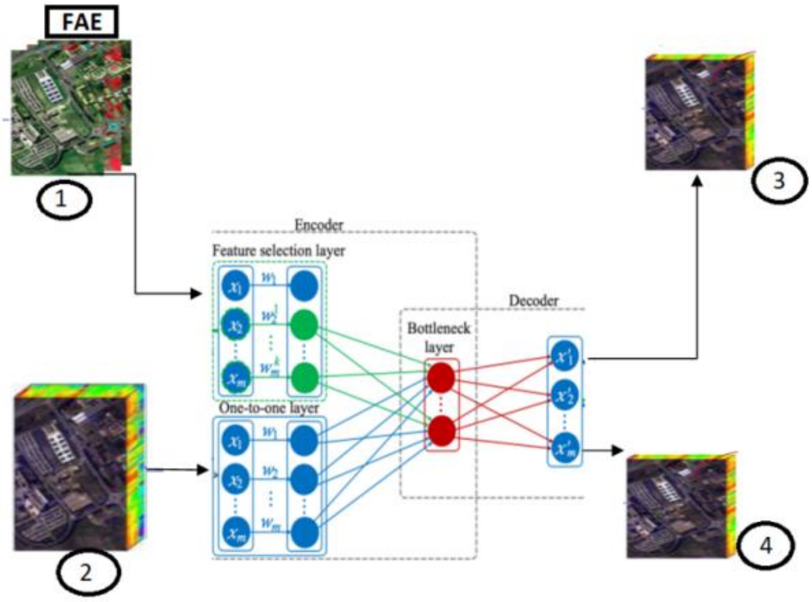

Figure 1. The architecture of FAE. The presented quantifies are: (1) feature selection result, (2) input, (3) reconstruction based on the selected features, (4) reconstruction from the one-to-one layer.

The operation of FAE is formalized as an optimization problem with two key components: The global reconstruction term minimizes the reconstruction error between the original HSI data X and the data reconstructed after passing through the encoder (g) and decoder (f) networks, considering the selected features represented by 'WI'. The diversity term is introduced to encourage the selected subset of features (WI) to be diverse and not highly correlated with each other. This term ensures that the chosen features effectively capture various aspects of the hyperspectral data.

$$min_{w,g,f}||\mathbf{X} - \mathbf{f}(\mathbf{g}(\mathbf{W_I})\ ||_F^2 + \boldsymbol{\lambda_1}||\mathbf{X} - \mathbf{f}(\mathbf{g}(\mathbf{W_I^{max\,k}})\ ||_F^2 + \boldsymbol{\lambda_2}||\mathbf{W_I}||_1,\ \mathbf{s.t.\,W_1} \geq \mathbf{0} \qquad \mathbf{(2)}$$

The overall objective function is balanced between these two terms, as shown in Equation (2), and is controlled by non-negative balancing parameters, $\lambda1$ and $\lambda2$. This approach is named FAE because of its intriguing characteristic: a small proportion of features selected in the second term can achieve performance similar to using the entire set of features in the first term when reconstructing the original hyperspectral data. This self-similarity trait becomes even more evident when FAE is applied to extract multiple feature subsets for different tasks. Firstly, FAE is utilized to perform feature selection on the hyperspectral data. FAE is tailored to select a subset of informative spectral bands from the original dataset while ensuring that the chosen features are diverse. This process aims to enhance the representativeness of the feature subset.

Once feature selection with FAE is completed, distinct supervised classification tasks are carried out. Supervised classification using ensemble learning algorithms, namely Random Forest, LightGBM (Gradient Boosting), XGBoost, and CatBoost, is also performed simultaneously. These classifiers are known for their robustness and ability to handle complex feature spaces. The selected features derived from FAE are used as inputs for these classifiers, which improves classification accuracy and interpretability.

This methodology enables a comprehensive evaluation of the effectiveness of FAE-based feature selection in supervised classification scenarios, contributing to a deeper understanding of hyperspectral data analysis techniques.

3. Experiments

3.1 Dataset Description

In this paper, the benchmarking dataset used is Pavia University. This data is commonly used in the HSI domain to assess and compare the performance of HSI processing and analysis algorithms.

3.2. Result and Discussion

In our study, the application of FAE for feature selection yielded notable improvements in classification performance. When compared to alternative feature selection methods, FAE consistently demonstrated superior results across various evaluation metrics, including accuracy, F1-score, recall, precision, and reconstruction error, which is measured in mean squared error (MSE) for evaluating the model.

The observed enhancements in classification accuracy and other performance metrics are underscored by the effectiveness of FAE in extracting the most relevant and diverse set of features from hyperspectral data. This robust feature selection process not only helps to reduce dimensionality but also ensures that the selected features retain critical information about the spectral content. The advantage of FAE over other feature selection techniques lies in its ability to strike a balance between preserving critical spectral information and promoting feature diversity. This characteristic makes FAE particularly well-suited for hyperspectral data, where a delicate balance between feature informativeness and redundancy is essential.

Table 1. Performance Accuracy Metric.

	Accuracy				
	UDFS	**MCFS**	**AE**	**PCA**	**FAE**
RF	0.48	0.80	0.85	0.79	0.85
LGBM	0.44	0.57	0.56	0.55	0.57
XGBOOST	0.55	0.81	0.87	0.80	0.85

CATBOOST	0.50	0.81	0.82	0.50	0.85

Table 2. Performance F1-Score Metric.

			F1-Score		
	UDFS	**MCFS**	**AE**	**PCA**	**FAE**
RF	0.26	0.87	0.86	0.82	0.90
LGBM	0.61	0.86	0.86	0.82	0.89
XGBOOST	0.29	0.82	0.82	0.83	0.88
CATBOOST	0.29	0.87	0.87	0.83	0.89

Table 3. Performance Recall Metric.

			recall		
	UDFS	**MCFS**	**AE**	**PCA**	**FAE**
RF	0.24	0.88	0.86	0.80	0.91
LGBM	0.00	0.84	0.80	0.77	0.89
XGBOOST	0.30	0.85	0.86	0.83	0.88
CATBOOST	0.29	0.86	0.86	0.83	0.88

4. Conclusion and Future Work

The results of supervised classification with various feature selection methods showed mixed outcomes. While some methods performed better than others, the classification results using feature selection by the Fractal Autoencoder (FAE) method were the most promising. This was done in an effort to minimize both time and costs associated with hyperspectral data processing and utilization, while still achieving satisfactory classification results. This choice not only streamlined the workflow but also reduced memory and computational requirements, making the overall process more efficient and cost-effective.

Reference

1. Wu, X., & Cheng, Q. (2021, May). Fractal autoencoders for feature selection. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 12, pp. 10370-10378).

2. Lu, Y., Cohen, I., Zhou, X. S., & Tian, Q. (2007, September). Feature selection using principal feature analysis. In Proceedings of the 15th ACM international conference on Multimedia (pp. 301-304).

3. Wah, Y. B., Ibrahim, N., Hamid, H. A., Abdul-Rahman, S., & Fong, S. (2018). Feature selection methods: Case of filter and wrapper approaches for maximising classification accuracy. Pertanika Journal of Science & Technology, 26(1).

4. Sousa, C. M., Cavalcante, A. B., Guilhon, D., & Barros, A. K. (2007). Image compression by redundancy reduction. In Independent Component Analysis and Signal

Separation: 7th International Conference, ICA 2007, London, UK, September 9-12, 2007. Proceedings 7 (pp. 422-429). Springer Berlin Heidelberg.

5. Balın, M. F., Abid, A., & Zou, J. (2019, May). Concrete autoencoders: Differentiable feature selection and reconstruction. In International conference on machine learning (pp. 444-453). PMLR.

6. Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., & Chanussot, J. (2013). Hyperspectral remote sensing data analysis and future challenges. IEEE Geoscience and remote sensing magazine, 1(2), 6-36.

7. Cai, D., Zhang, C., & He, X. (2010, July). Unsupervised feature selection for multi-cluster data. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 333-342).