UNSUPERVISED CLUSTER-WISE HYPERSPECTRAL BAND SELECTION USING CLUSTERING SILHOUETTE INFORMATION

Mateus Habermann[1], *Marielcio Gonçalves Lacerda*[2]

Institute for Advanced Studies, São José dos Campos - SP, [1]habermannmh@fab.mil.br, [2]marielciomgl@fab.mil.br

RESUMO

Devido às suas várias bandas, uma imagem hiperspectral pode prover informações detalhadas sobre um cenário, porém a alta dimensão desses dados pode causar o overfitting do classificador, além de causar a maldição de dimensionalidade. Seleção de bandas pode amenizar esses problemas. Assim, este artigo aborda essa questão ao incrementar e melhorar um algoritmo não supervisionado sobre seleção de bandas que foi recentemente publicado. Especificamente, o uso de informação de Silhouette é usado para se determinar automaticamente a quantidade de clusters —e iterações— desse método. Essa quantidade de clusters tem sido até então considerada um parâmetro informado pelo usuário. Experimentos mostram que a melhoria proposta atinge resultados similares aos do método original, com a vantagem de ser independente da ingerência do usuário.

Palavras-chave – seleção de bandas, não supervisionado, informação Silhouette, imagem hiperespectral, maldição de dimensionalidade.

ABSTRACT

Due to its multiple bands, a hyperspectral image may provide fine details about a scene, but the high dimensionality may cause both overfitting of the classifier and the Hughes phenomenon. Band selection can alleviate this problem. So, this paper addresses this issue by improving a recently published unsupervised band selection algorithm. Specifically, we propose the use of Silhouette information in order to determine the number of clusters —and iterations— of the method. This number had been so far determined by the user. Experiments show that the improved algorithm achieves similar results to those of the original method, with the advantage of being independent of user's knowledge.

Key words – *band selection, unsupervised, Silhouette information, hyperspectral image, Hughes phenomenon.*

1. INTRODUCTION

In data classification problems, the classes separation in the feature space plays a key role [1]. The more separable the classes are, easier is the job of the classifier.

Concerning the classification of hyperspectral images (HSI), a good selection of bands can provide a wider class separation [2], attenuate the Hughes phenomenon [3] and avoid the overfitting of the classifier [4–6].

In order to tackle this issue, one could resort to feature extraction, which proceeds to a combination of the original bands. However it is not a plausible method for hyperspectral data dimensionality reduction, because the resulting features lose the physical information [7].

Band selection (BS) also addresses the same high dimensionality issues, but has the advantage of keeping the original information of the spectral bands [7, 8].

Unsupervised BS algorithms follow either a ranking-based [9] or a clustering-based approach [10]. Ranking-based frameworks use a specific criterion to sort the spectral bands. Band selection algorithms based on clustering seek to find the most representative bands of each cluster, and make the correlation amongst bands smaller [11].

In the BS literature, clustering is normally used to form clusters of spectral bands [10, 12–16]. By doing so, data structural information is not considered. When band selection is performed for classification purposes, an analysis based only on the best representative bands —without considering class separation— is not of great importance.

In order to address those aspects, [17] presented a band selection algorithm which sought to maximize the distance amongst classes in the feature space. This approach uses clustering on the data set, and the number of clusters is defined by the user. But it is clearly a deficiency of the method, because one must rely on the experience and knowledge of the user in order to determine the number of clusters in the data. So, to solve this problem, this paper presents an improvement of the algorithm by devising a method to automate the choice for the parameter *number of clusters*.

So, in the proposed approach, the clustering operations take place with different number of clusters. For each clustering operation, the Silhouette information is summed for all data entries. At the end of the operations, the one with the biggest sum defines the number of clusters for that data set.

The contribution of this paper is this:

- The automation of the parameter *number of clusters*, which had been previously defined by the user.

In Section 2, the improvement on the method and the image used in this work are presented. In Section 3, the results of the two versions of the method are compared by the classification of a hyperspectral image. In Section 4 an analysis of the results is given. Finally, in Section 5 we conclude this paper.

2. MATERIAL AND METHODS

The band selection method to be improved is given by Algorithm 1, where $\mathbf{C}_0$ is the whole data set; k is the user-defined number of clusters; $\mathbf{S}$ is the subset to keep the selected bands; $\mathbf{D}$ is the subset to keep the discarded bands; s is the number of bands to be selected; and w are the parameters of a hyperplane.

Obviously, it is important to refer to the original paper [17] to understand the method.

We propose an improvement on the Line 1 of the Algorithm 1, by turning the parameter k from user-defined to automated. For this, we resort to Silhouette clustering [18].

Algorithm 1 Band selection algorithm to be improved

1: **Input:** Data set $\mathbf{C}_0$; user-defined number k of clusters
2: $\mathbf{S} = \emptyset$ ▷ Subset of selected bands
3: $\mathbf{D} = \emptyset$ ▷ Subset of bands to be discarded
4: Proceed to K-Means clustering (*cosine* distance) of $\mathbf{C}_0$ into k clusters $\mathbf{C}_i$
5: **for** i = 1 : k **do**
6: Proceed to a binary classification between clusters $\mathbf{C}_i$ and $\mathbf{C}_0 \setminus \mathbf{C}_i$ (*one-versus-all*) using a single-layer neural net
7: Select the $4(s/k) \in \mathbb{N}$ bands related to the biggest separating hyperplane parameters $|w|$
8: Proceed to the band selection fine-tuning
9: Update subset of selected bands $\mathbf{S}$
10: Update subset $\mathbf{D}$
11: Update the whole data set
12: **Return: S**

2.1. Silhouette clustering information

Let's assume the data set have been clustered. For a data point $x_i \in C_I$, let

$$a(x_i) = \frac{1}{|C_I| - 1} \sum_{x_j \in C_I, x_i \neq x_j} d(x_i, x_j)$$

be the mean distance between x_i and all other data points in the same cluster, where C_I is the cluster cardinality, and $d(x_i, x_j)$ is the distance between x_i and x_j. a(i) can be seen as a measure of how well x_i is assigned to its cluster —the smaller the value, the better the assignment.

Let us also define the mean dissimilarity of x_i as the mean distance between it and the elements in a cluster $C_J \neq C_I$. Given $x_i \in C_I$, we have

$$b(x_i) = \min_{J \neq I} \frac{1}{|C_J|} \sum_{x_j \in C_J} d(x_i, x_j)$$

to be the smallest mean distance of x_i to all data entries in any other cluster. The cluster with this smallest mean dissimilarity is considered the *neighboring cluster*.

Now it is possible to define the silhouette value for x_i as

$$s(x_i) = \frac{b(x_i) - a(x_i)}{max(a(x_i), b(x_i))}, \quad (1)$$

if $|C_I| > 1$, and

$$s(x_i) = 0, \text{ if } |C_I| = 1.$$

At the end of this procedure, each data entry has its own s value, which is dubbed *silhouette*.

By the definitions above, we see that $-1 \leq s(x_i) \leq 1$. Besides, if $s(x_i) \approx 1$ it means that $a(x_i) \ll b(x_j)$ and this indicates that x_i is similar to the elements of its own clusters and dissimilar to the others.

Thus, one can cluster a data set using different number of clusters. The configuration with the biggest summation of s is considered the best one. In this paper, we proceed to clustering with different number of clusters. And we adopt the setting whose summation s value is the biggest.

2.2. Data set

In this paper, for the sake of results comparison, we use the Botswana HSI. This image is composed of 242 spectral bands and has 14 classes. See further details about this image at [19].

3. RESULTS

This paper aims at improving [17] by letting the parameter k user-free.

Given the Botswana data set, we proceed to different clustering operations, varying the number of clusters. For each clustering, the s value for each data point is calculated according to Equation 1, and the summation is stored. At the end, we set the number of clusters the one with the biggest s summation. In this case, the number of clusters with the biggest s summation is 4. This fact is depicted by Figure 1.

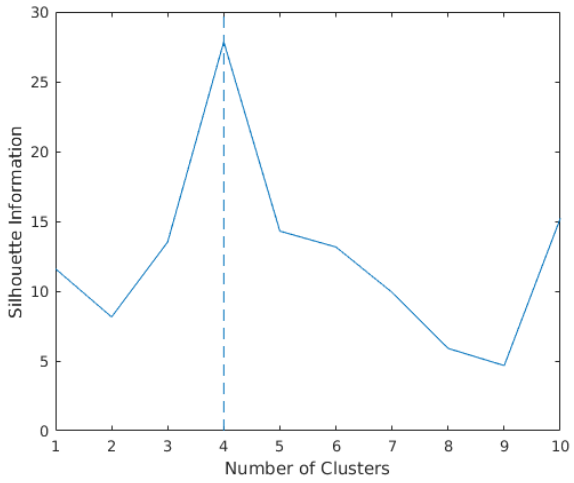

Figura 1: Silhouette information for different number of clusters. The dashed line indicates that the biggest score is attained with four clusters.

Thus, the method can be updated as depicted in Algorithm 2, in Line 2.

Using Algorithm 2 one can select the hyperspectral bands. After this, we proceed to supervised classification using the select bands as input. For the classification we use k-Nearest Neighbors (KNN), Classification and Regression Trees (CART) and Support Vector Machine (SVM) [1]. Table 1 shows the classification results between the improved method (CW_S) and the original method (CW). Four different number of bands are selected, for the sake of comparison.

From Table 1, the proposed method outperforms its original version in 7 out of 12 experiments. Figures 2, 3 and

Algorithm 2 Improved band selection algorithm

1: **Input:** Data set $\mathbf{C}_0$
2: Define number k according to Equation 1
3: $\mathbf{S} = \emptyset$ ▷ Subset of selected bands
4: $\mathbf{D} = \emptyset$ ▷ Subset of bands to be discarded
5: Proceed to K-Means clustering (*cosine* distance) of $\mathbf{C}_0$ into k clusters $\mathbf{C}_i$
6: **for** i = 1 : k **do**
7: Proceed to a binary classification between clusters $\mathbf{C}_i$ and $\mathbf{C}_0 \setminus \mathbf{C}_i$ (*one-versus-all*) using a single-layer neural net
8: Select the $4(s/k) \in \mathbb{N}$ bands related to the biggest separating hyperplane parameters $|w|$
9: Proceed to the band selection fine-tuning
10: Update subset of selected bands $\mathbf{S}$
11: Update subset $\mathbf{D}$
12: Update the whole data set
13: **Return: S**

4 show the results in an illustrative manner.

Method	10 bands	20 bands	30 bands	40 bands
		KNN Classifier		
CW_S	89.49±0.69	91.18±0.97	90.33±0.96	91.19±1.17
CW	90.30±0.94	90.48±0.86	91.38±1.04	90.57±0.69
		CART Classifier		
CW_S	83.30±1.47	85.55±1.49	83.85±1.03	85.04±1.59
CW	83.21±0.72	85.05±1.16	85.48±1.27	85.69±0.70
		SVM Classifier		
CW_S	90.94±0.70	93.58±0.70	94.25±0.60	94.94±0.84
CW	90.91±0.98	93.41±0.94	94.61±0.84	94.13±0.77

Tabela 1: Mean classifiers accuracies and standard deviation in %.

4. DISCUSSION

The proposed improvement on [17] lies in Line 2 of Algorithm 2.

We do not rely on the expertise of the user to define k anymore, it is now done automatically.

Running in MatLab using a processor Intel Core i7-5500U CPU @ 2.40GHz x 4, Line 2 of Algorithm 2 takes about 31 seconds, plus the mean time the original algorithm takes to process (about 43 seconds). It means that the improvement almost doubles the mean processing time. On the other hand, this method is supposed to run off-line, so $31 + 43 = 74$ seconds is not something we should worry about. Anyway, by using more powerful computers this processing time will be shorter.

Considering all the experiments in Table 1, CW_S gets the best results in 7 out of 12 experiments, and has the mean accuracy of $89.47 \pm 0.49\%$, whereas CW achieves $89.60 \pm 0.58\%$. That is, they are statistically equal. Therefore, one could say that turning k an automated parameter brings no loss to the method, and can, consequently, be used.

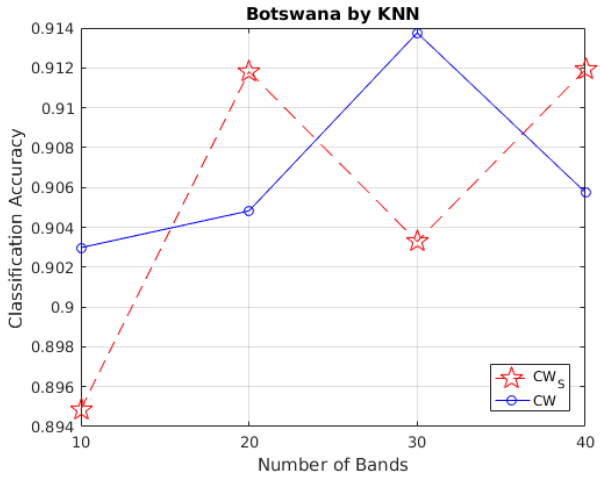

Figura 2: As legendas das figuras (e também tabelas) devem ser formatadas com fonte 9 pontos, em negrito, e centralizadas na coluna.

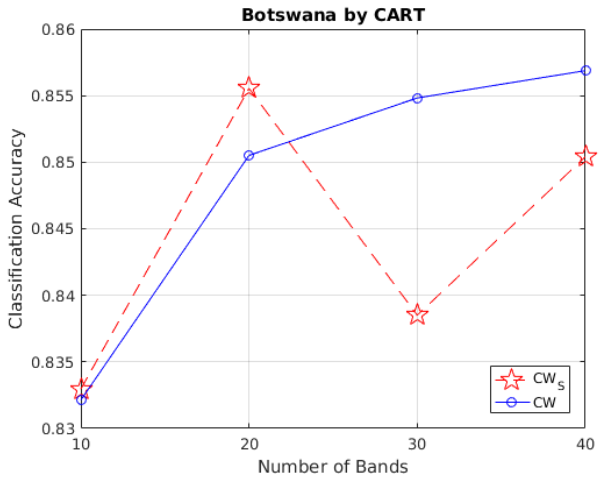

Figura 3: As legendas das figuras (e também tabelas) devem ser formatadas com fonte 9 pontos, em negrito, e centralizadas na coluna.

5. CONCLUSION

This paper proposed an improvement on a recently published band selection paper.

More precisely, the original method considered the parameter k to be user-defined. The problem lies in the fact that one must rely on an expert to set this parameter accordingly. This person is not always available.

So, this paper used Silhouette clustering information in order to define automatically the k number of clusters to be formed by the algorithm.

Experimental results using a hyperspectral image showed that the proposed method and its original version have similar accuracy classification results. Thus one can conclude that the proposed improvement is capable of automating a input parameter and keeping the same good results.

For further works, we suggest the analysis of other methods for the definition of an optimum number of clusters in a data

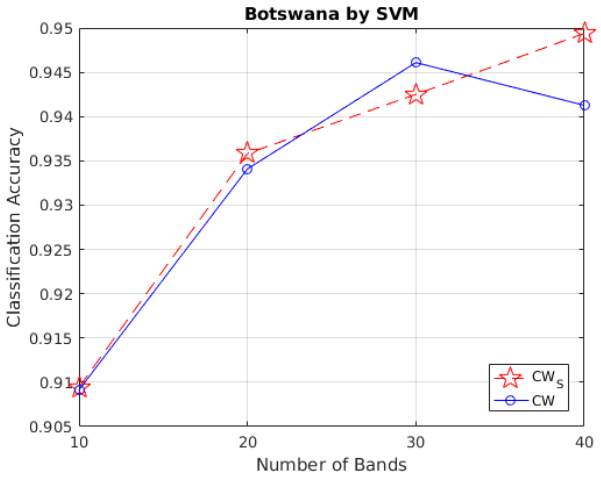

Figura 4: As legendas das figuras (e também tabelas) devem ser formatadas com fonte 9 pontos, em negrito, e centralizadas na coluna.

set.

6. REFERENCES

[1] Sergios. Theodoridis and Konstantinos. Koutroumbas. *Pattern recognition*. Elsevier, 4th ed. edition, 2009.

[2] Mateus Habermann, Vincent Frémont, and Elcio Hideiti Shiguemori. Supervised band selection in hyperspectral images using single-layer neural networks. *International Journal of Remote Sensing*, 40(10):3900–3926, 2019.

[3] Michael E. Houle, Hans-Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek. Can shared-neighbor distances defeat the curse of dimensionality? In Michael Gertz and Bertram Ludäscher, editors, *Scientific and Statistical Database Management*, pages 482–500, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[4] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. Wiley, New York, 2 edition, 2001.

[5] Simon S. Haykin. *Neural networks and learning machines*. Pearson Education, Upper Saddle River, NJ, third edition, 2009.

[6] Ludmila I. Kuncheva, Clare E. Matthews, Álvar Arnaiz-González, and Juan J. Rodríguez. Feature selection from high-dimensional data with very low sample size: A cautionary tale, 2020.

[7] X. Shang, M. Song, Y. Wang, C. Yu, H. Yu, F. Li, and C. Chang. Target-constrained interference-minimized band selection for hyperspectral target detection. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–21, 2020.

[8] X. Ding, H. Li, J. Yang, P. Dale, X. Chen, C. Jiang, and S. Zhang. An improved ant colony algorithm for optimized band selection of hyperspectral remotely sensed imagery. *IEEE Access*, 8:25789–25799, 2020.

[9] B. Xu, X. Li, W. Hou, Y. Wang, and Y. Wei. A similarity-based ranking method for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–15, 2021.

[10] A. Datta, S. Ghosh, and A. Ghosh. Combination of clustering and ranking techniques for unsupervised band selection of hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2814–2823, 2015.

[11] W. Xie, J. Lei, J. Yang, Y. Li, Q. Du, and Z. Li. Deep latent spectral representation learning-based hyperspectral band selection for target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 58(3):2015–2026, 2020.

[12] F. Zhang, Q. Wang, and X. Li. Hyperspectral image band selection via global optimal clustering. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1–4, 2017.

[13] G. Tang, S. Jia, and J. Li. An enhanced density peak-based clustering approach for hyperspectral band selection. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1116–1119, 2015.

[14] W. Sun, J. Peng, G. Yang, and Q. Du. Correntropy-based sparse spectral clustering for hyperspectral band selection. *IEEE Geoscience and Remote Sensing Letters*, 17(3):484–488, 2020.

[15] V. Kumar, J. Hahn, and A. M. Zoubir. Band selection for hyperspectral images based on self-tuning spectral clustering. In *21st European Signal Processing Conference (EUSIPCO 2013)*, pages 1–5, 2013.

[16] M. Bevilacqua and Y. Berthoumieu. Multiple-feature kernel-based probabilistic clustering for unsupervised band selection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6675–6689, 2019.

[17] Mateus Habermann, Elcio Hideiti Shiguemori, and Vincent Frémont. Unsupervised cluster-wise hyperspectral band selection for classification. *Remote Sensing*, 14(21), 2022.

[18] Guo Jin-Heng, Lin Jia-Xiang, Zhang Zhen-Chang, and Ling Han-Yu. Cdbscan: Density clustering based on silhouette coefficient constraints. In *2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI)*, pages 600–605, 2022.

[19] M Graña and B Veganzons, MA Ayerdi. Hyperspectral remote sensing scenes. `http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes`.