Divisive Hierarchical Clustering for PolSAR images

Carvalho, N.[1], Sant'anna, S.[2], Bins, L[2].

[1]Programa de Pós-Graduação em Computação Aplicada - CAP
[2]Divisão de Processamento de Imagens – DPI
Instituto Nacional de Pesquisas Espaciais – INPE
naiallen@yahoo.com.br, leonardo@dpi.inpe.br, sidnei@dpi.inpe.br

Introduction

Hierarchical clustering algorithms perform the data clustering by developing a binary tree-based data structure called dendrogram (Figure 1). Those algorithms can be grouped into two techniques [1]:

- **Agglomerative or "bottom-up" approach, which is based on clustering merge;**
- **Divisive or "top-down" approach, which is based on clustering split.**

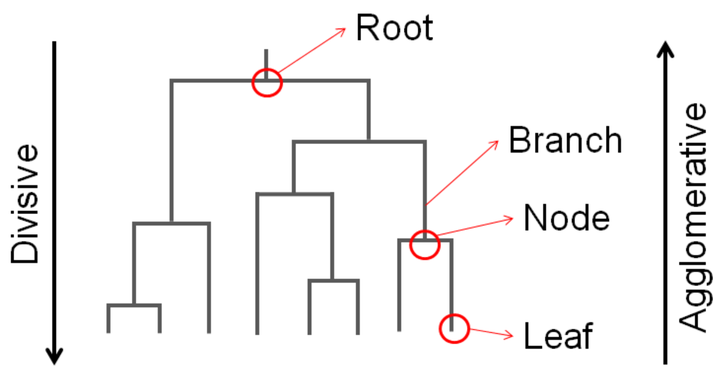

Figure 1: Hierarchical Clustering dendrogram.

In this work we propose a novel divisive hierarchical algorithm for clustering PolSAR (Polarimetric Synthetic Aperture Radar) images, using a stochastic approach.

Proposed Algorithm

The inspiration for this algorithm came from the Bisecting K-Means clustering Algorithm [2]. However, in this work, we propose to use the PDDP (Principal Direction Divisive Partitioning) algorithm in order to determine the initial centroids, instead of randomly chose it. We divide the dataset into two dummy clusters using the plan over the principal direction as the frontier, the mean value of each of those clusters will be the initial centroid. After that, the bisecting is performed using the stochastic distance of Bhattacharya as a metric.

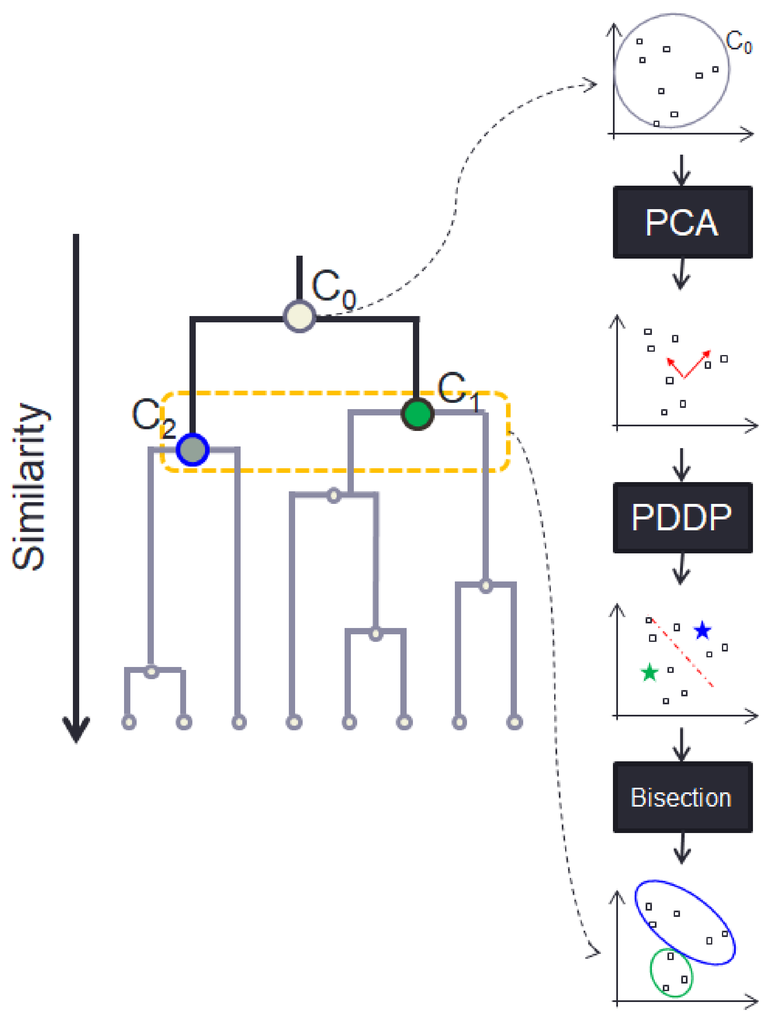

Figure 2: Proposed divisive hierarchical clustering algorithm.

The PolSAR image is widely known as following the Wishart distribution, therefore the Bhattacharya distance between two Wishart is defined as:

$$d_{\mathrm{WB}}(\Sigma_1,\Sigma_2) = L\left[\frac{\log|\Sigma_1| + \log|\Sigma_2|}{2} - \log\left|\left(\frac{\Sigma_1^{-1}+\Sigma_2^{-1}}{2}\right)^{-1}\right|\right]$$

where Σ is the covariance matrix and L is the image number of looks.

Results

We applied the classification algorithms above-cited on a simulated PolSAR image to generate the dendrogram showed in Figure 3. Figure 4 shows a simulated PolSAR image and its classification using the proposed algorithm. Since the image is simulated, we used knowns samples from it to compute an overall accuracy. The classification achieved 82.54% of overall accuracy, mainly due to the wrong split of classes 6, 7 and 8, which, in reality, form just one class.

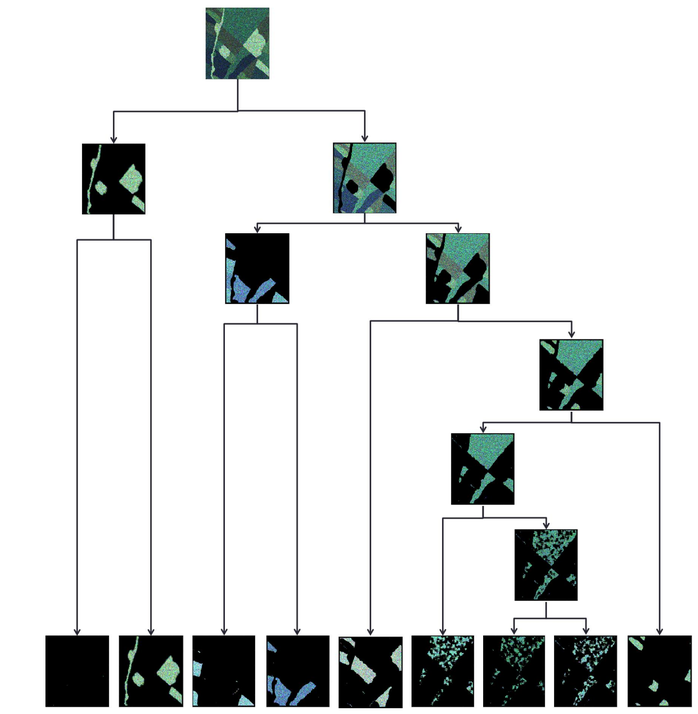

Figure 3: Classification dendrogram .

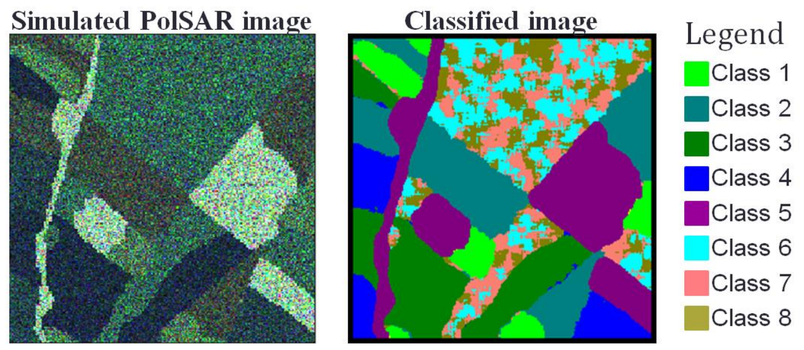

Figure 4: Simulated PolSAR image and Classified image.

Conclusion

The results are promising, and the development is ongoing. Besides that, a good feature about this algorithm is: once the dendrogram is constructed, the number of clusters can be automatically chosen by splitting the tree at different levels to obtain different clustering solutions for the same dataset without rerunning the clustering algorithm again.

Agradecimentos

Os autores agradecem o apoio financeiro da CAPES, CNPq, FAPESP e FINEP.

Referências

[1] S. Bandyopadhyay and S. Saha. Unsupervised classification: similarity measures, classical and metaheuristic approaches, and applications. Springer Science & Business Media, 2012.

[2] R. Patil and A. Khan. Bisecting K-means for clustering web log data. International Journal of Computer Applications 116.19 (2015).