Clustering Remotely Sensed Data with K-Means and Bat Algorithm

Lamia F. H. CHAOUCHE RAMDANE[1*], Habib MAHI[2]

[1]Laboratory of Research in Computer Science, Faculty of Sciences, University of Tlemcen, Algeria
[2]Earth Observation Division, Centre of Space Techniques (CTS), Arzew, Algeria
*Corresponding author's e-mail: chaouche_lamia@yahoo.fr

ABSTRACT

K-Means is an efficient clustering method commonly employed in data analysis, particularly in the field of remote sensing. However, selecting the right number of clusters (k) can be challenging. To tackle this challenge, we investigate the effectiveness of K-Means combined with the Bat Algorithm (KMBA), which leverages bat-inspired techniques. The results show promise for unsupervised classification using composite data images and remote sensing data.

Keywords: Clustering; KMBA; Validity index

1 Introduction

Unsupervised classification in remote sensing involves clustering images into spectral classes. Common algorithms like K-means, ISODATA, and SOM have limitations [1]. Researchers turn to metaheuristic optimization methods [2], [4], such as the bat algorithm, which this study explores. It uses the bat optimization algorithm and proposes the K means and Bat Algorithm (KMBA) for efficient clustering of composite data images and remote sensing data. KMBA combines K-means and the Bat Algorithm, eliminating the need to pre-specify the number of clusters and cluster centers. The study employs the Davies-Bouldin index (DB) as an objective function for evaluation. Section 2 covers echolocation and the bat algorithm, Section 3 presents experimental data and results, and Section 4 concludes.

2 Methodology

The KMBA (K-Means combined with Bat Algorithm) clustering algorithm, introduced by Komarasamy and Wahi in 2012 [5], combines the principles of K-Means [6] and the Bat Algorithm [7]. Each bat in the algorithm plays a dual role in initializing cluster locations. The first task involves using a discrete Particle Swarm Optimization (PSO) approach, inspired by the Cooperative Particle Swarm Optimizer (CPSO) [8], to determine the optimal number of clusters. The second task employs the K-Means algorithm to identify the best cluster center set based on the assigned number of clusters. This combination results in efficient clustering without the need for users to pre-specify the number of clusters and cluster centers, with the DB validity index [9] serving as the objective function for evaluation.

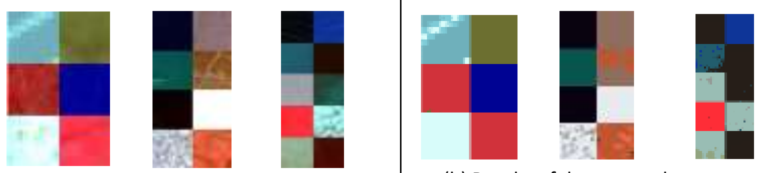
3 Results and Discussion

The KMBA algorithm's performance relies on several parameters, including pulse rate, intensity, frequency range, number of iterations, and bat population. The frequency is typically within the interval [fmin, fmax], with a wide range for thorough investigation. In this case, we chose [0, 500] with f = 500, representing the maximum bat frequency. The pulse rate and intensity were fixed (Ai = ri = 0.25, $\forall$ I = 1, n). Varying the bat population size yielded different index values for xi, with the figure displaying the best values and their corresponding number of clusters (K) for composite images.

We propose applying this approach to three synthetic satellite images, each containing 6, 8, and 10 clusters, as illustrated in Figure 1(a). These images were created by manually selecting distinct samples from a satellite image, resizing them, and merging them into a composite image. This method allows for the generation of synthetic satellite images with predefined classes, facilitating testing. In contrast to the results obtained with

©2024 Copyright held by the author(s). Published by AIJR Publisher in "Abstracts of the First International Conference on Advances in Electrical and Computer Engineering 2023" (ICAECE'2023) 15-16 May 2023. Organized by the Faculty of Science and Technology, Department of Electrical Engineering, University of Echahid Cheikh Larbi Tebessi, Tebessa-Algeria.
DOI: 10.21467/abstracts.163 ISBN: 978-81-965621-5-1 (eBook)

the traditional K-Means algorithm (where the number of clusters is 7, 7, and 8, corresponding to images 1, 2, and 3, respectively), Figure 1 (b) demonstrates that this approach yields the precise number of clusters with a minimal DB index for the resulting image, highlighting the effectiveness of the KMBA algorithm.

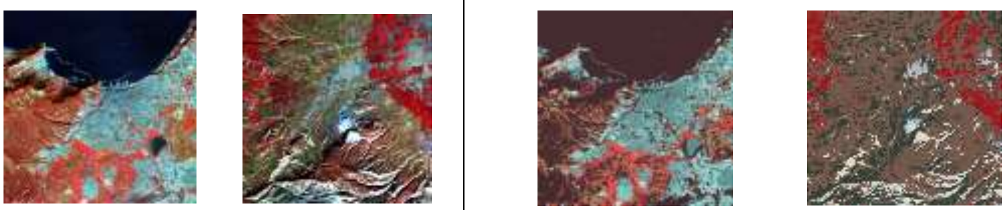


(a) Composite images used

(b) Results of the approach on composite images

Figure 1: *Composite images and results*

We suggest using the KMBA algorithm on multispectral images from Landsat (30m) and Spot (20m) satellites that depict the city of Oran in Algeria. This region has a diverse range of classes such as urban, soil, water, vegetation, forest, and cultivation, making it relatively complex, as illustrated in Figure 2 (a).
In urban or peri-urban areas, spectral variability is high, leading to more clusters compared to homogeneous areas. The bat algorithm offers optimized clustering, determining the ideal class count by minimizing the DB index (Figure 2, b). However, bat population affects results, potentially causing class confusion and missing regions with more bats. Execution time is critical; fewer bats reduce solution diversity, raising the DB index and decreasing optimal clustering potential.



(a) satellite images used (Landsat and Spot)

(b) Results of the approach on satellite images (11 cluster and 8 cluster)

Figure 2: *Remotely Sensed Data and results*

4 Conclusion

We apply metaheuristics to tackle image clustering, offering optimization benefits. We have adopted the bat algorithm, mimicking bats' echolocation, which yields an optimal cluster count and the best-scored image classification via an objective function. We selected the DB index as it is widely used for clustering satellite images. Our tests demonstrate the efficiency and effectiveness of this metaheuristic approach. Future research will explore the use of other validity indices and additional metaheuristic methods.

How to Cite

L. F. H. C. RAMDANE, H. MAHI, "Clustering Remotely Sensed Data with K-Means and Bat Algorithm", *AIJR Abstracts*, pp. 128–130, Feb. 2024.

References

[1] L.H.F. Chaouche Ramdane, H. Mahi, M. El Habib Daho, and M.A. Lazouni, "Multiple classifier system for remotely sensed data clustering", IET Image Process. 1(16), 2022, pages 252-260.
[2] Y. Kumar & A. Kaur, "Variants of bat algorithm for solving partitional clustering problems", Engineering with Computers, 2021.
[3] L. Zhu & J. Wang. "Data Clustering Method Based on Bat Algorithm and Parameters Optimization". Engineering Letters, 2019.
[4] Rozlini & Samsudin, "An Optimized Discretization Approach using k-Means Bat Algorithm". Turkish Journal of Computer and Mathematics Education, 2021, Vol.12 No.3, 1842-1851
[5] Komarasamy & Wahi, "An optimized K-means clustering technique using bat algorithm", European J. Scientific Research, 2012.

ISBN: 978-81-965621-5-1
Series: AIJR Abstracts
DOI: 10.21467/abstracts.163

[6] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations", Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1967, 1:281-297.

[7] X. Yang, "A New Metaheuristic Bat-Inspired Algorithm", University of Cambridge, Springer, 2010, Vol. 1, 1004.4170.

[8] F. Vanden Bergh, and A. P. Engelbrecht, "A Cooperative Approach to Particle Swarm Optimization", IEEE Transactions on Evolutionary Computation, 8(3), 2004, 225–239.

[9] Q. Zhao, "Cluster Validity in Clustering Methods", Publications of the University of Eastern Finland Dissertations. No 77, 2012.

ISBN: 978-81-965621-5-1

Series: AIJR Abstracts

DOI: 10.21467/abstracts.163