



APLICAÇÃO DO ALGORITMO K-MEANS NO AGRUPAMENTO DE IMAGENS DE SATÉLITE PARA IDENTIFICAÇÃO DE SELEÇÕES EM CLUSTERS

João G. M. SILVA¹; João P. do N. SILVA²; Diego SAQUI³; Allan A. PEREIRA⁴

RESUMO

Este artigo aborda uma aplicação de inteligência artificial em imagens de satélite, concentrando-se particularmente no algoritmo K-Means para agrupar seleções em clusters. O estudo destaca a relevância crucial dessa abordagem na análise e interpretação de vastos volumes de dados provenientes de imagens espaciais, bem como seu potencial promissor para promover avanços significativos em diversas áreas de aplicação, como monitoramento ambiental, planejamento urbano e agricultura de precisão. Ao examinar as capacidades do algoritmo K-Means nesse contexto, o artigo visa proporcionar uma compreensão dos benefícios e das implicações práticas associadas à utilização dessa técnica inovadora no domínio das imagens de satélite.

Palavras-chave:

RGB; NDVI; Inteligência Artificial; Clusterização; Earth Engine.

1. INTRODUÇÃO

Imagens de satélite fornecem uma rica fonte de informações sobre a Terra, permitindo a obtenção de dados geoespaciais em larga escala. No entanto, a interpretação manual dessas imagens pode ser um processo demorado e suscetível a erros. Nesse contexto, a aplicação de técnicas de Inteligência Artificial, como o algoritmo K-Means, oferece uma abordagem promissora para a análise automatizada e a clusterização eficiente de seleções em imagens de satélite.

Ao utilizar o algoritmo K-Means, é possível agrupar pixels de uma imagem de satélite com base em suas características espectrais. Esses grupos representam diferentes categorias presentes na imagem, como áreas urbanas, corpos d'água, vegetação e solo exposto. Esse agrupamento automatizado permite uma análise mais rápida e precisa de grandes conjuntos de dados de imagens, fornecendo informações valiosas para a tomada de decisões em várias áreas de aplicação.

No entanto, o agrupamento de pixels de imagens de satélite com o algoritmo K-Means apresenta desafios específicos. As imagens de satélite possuem alta dimensionalidade e uma ampla gama de valores espectrais, o que pode resultar em um agrupamento impreciso se não forem considerados cuidadosamente os parâmetros e o pré-processamento adequado. Além disso, a interpretação dos grupos gerados pelo K-Means requer um conhecimento prévio do ambiente e das características das áreas a serem identificadas. Portanto, é fundamental explorar técnicas de

¹ Discente do Bacharelado em Ciência da Computação, IFSULDEMINAS – Campus Muzambinho. E-mail: jgmsgabriel@gmail.com

² Coautor, IFSULDEMINAS – Campus Muzambinho. E-mail: jpns1118 @ gmail.com

³ Orientador, IFSULDEMINAS – Campus Muzambinho. E-mail: diego.saqui@muz.ifsuldeminas.edu.br

⁴ Coorientador, IFSULDEMINAS – Campus Muzambinho. E-mail: allan.pereira@ifsuldeminas.edu.br

pré-processamento e estratégias de avaliação para melhorar a precisão e a eficiência do processo do agrupamento utilizando o algoritmo K-Means em imagens de satélite.

2. FUNDAMENTAÇÃO TEÓRICA

As imagens de satélite são uma valiosa fonte de dados geoespaciais que capturam informações detalhadas da superfície terrestre. Obtidas por sensores em satélites, elas abrangem diferentes bandas do espectro eletromagnético, permitindo a detecção de características como vegetação, água, solo e áreas urbanas.

A interpretação manual dessas imagens é complexa e demorada, tornando essencial o uso de técnicas de processamento de imagens e inteligência artificial para automatizar a análise e extração de informações relevantes. Algoritmos de aprendizado de máquina e visão computacional aplicados às imagens de satélite possibilitam a detecção de padrões, segmentação de áreas de interesse e classificação de objetos, fornecendo informações valiosas para aplicações como monitoramento ambiental, planejamento urbano e avaliação de recursos naturais.

O algoritmo K-Means é um método amplamente utilizado para a clusterização de dados. Ele busca identificar grupos (*clusters*) nos dados, atribuindo pontos ao cluster mais próximo com base em suas características. No contexto de imagens de satélite, como explica USMAN (Babawuro), o K-Means pode ser aplicado considerando as propriedades espectrais dos pixels, como valores de intensidade em diferentes bandas do espectro eletromagnético.

A aplicação do K-Means em imagens de satélite requer a definição de parâmetros adequados, como o número de *clusters* desejados. É importante considerar a dimensionalidade e a escala das características espectrais para obter resultados precisos. Além disso, técnicas de pré-processamento, como a redução de dimensionalidade e a normalização de dados, podem ser aplicadas para melhorar a eficiência e a qualidade do agrupamento.

Em suma, o uso do algoritmo K-Means em imagens de satélite apresenta um grande potencial para o agrupamento. Essa abordagem permite uma análise mais precisa e rápida de grandes volumes de dados de imagens espaciais, fornecendo informações valiosas para diversas aplicações.

3. MATERIAL E MÉTODOS

Para realizar as análises neste estudo, foi utilizado o ambiente de programação Python juntamente com o serviço do Google Colab⁵, uma plataforma baseada em nuvem que permite a execução de código Python de forma interativa. O conjunto de dados Harmonized Sentinel-2 MSI:

⁵ Disponível em:

<https://colab.research.google.com/drive/1z9bi5_MnSJUFZuvobg5ficHCw1K47GnS?usp=sharing>. Último acesso em 30/06/2023

MultiSpectral Instrument, Level-2A⁶ utilizado neste trabalho foi carregado no ambiente do Google Colab, com a integração com o Google Earth Engine⁷ e as bibliotecas Python como Geemap⁸, Pycrs⁹ e Pysnp¹⁰ para a manipulação e processamento dos dados. No conjunto de dados foram utilizados 4 atributos, sendo eles as bandas B2, B3 e B4, e o índice NDVI gerado a partir da banda infravermelho próximo B8. Foram realizadas etapas de pré-processamento, como definir um intervalo de tempo, no qual foi considerado o ano de 2022, selecionar uma região de interesse de estudo, que foi definida como a área de Muzambinho e filtrar a imagem com menor percentual de nuvens. Além disso, para a aplicação do algoritmo K-Means, foi utilizada a biblioteca Earth Engine¹¹, que disponibiliza uma implementação eficiente e de fácil utilização dessa técnica de clusterização.

Para avaliar a vegetação presente na imagem, foi calculado o Índice de Vegetação por Diferença Normalizada (NDVI, do inglês Normalized Difference Vegetation Index). E por fim, utilizou-se o algoritmo K-Means, uma técnica de aprendizado não supervisionado, para realizar o agrupamento de imagens. O algoritmo foi aplicado, definindo o número de clusters como 05 (cinco), nas duas imagens, sendo uma utilizando as bandas B2 (Banda azul), B3 (Banda verde) e B4 (Banda vermelha), e a outra utilizando o índice NDVI.

4. RESULTADOS E DISCUSSÃO

Ao aplicar o algoritmo K-Means nas imagens de satélite, tanto na imagem RGB quanto na imagem do NDVI, foi obtido resultados significativos no agrupamento dos segmentos.

Na imagem RGB, os resultados do K-Means mostraram uma separação boa dos diferentes elementos presentes na imagem. No entanto, a distinção entre áreas vegetadas e áreas não vegetadas não foi tão clara quanto o esperado, conforme indica a Figura 1. Isso é atribuído à similaridade de cores entre certas áreas de vegetação e outros elementos, como telhados de casas. Por outro lado, ao aplicar o algoritmo K-Means na imagem do NDVI, os resultados foram mais precisos e distintos. Com essa informação, os clusters gerados pelo K-Means agruparam melhor as áreas de vegetação, facilitando a identificação de regiões verdes na imagem. Além disso, os clusters obtidos no NDVI permitiram uma melhor distinção entre os diferentes tipos de cobertura terrestre. Segmentos de asfalto e áreas descobertas foram separados de forma mais precisa.

⁶ Disponível em:

<https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED>. Último acesso em 30/06/2023

⁷ Disponível em: <<https://earthengine.google.com/>>. Último acesso em 30/06/2023

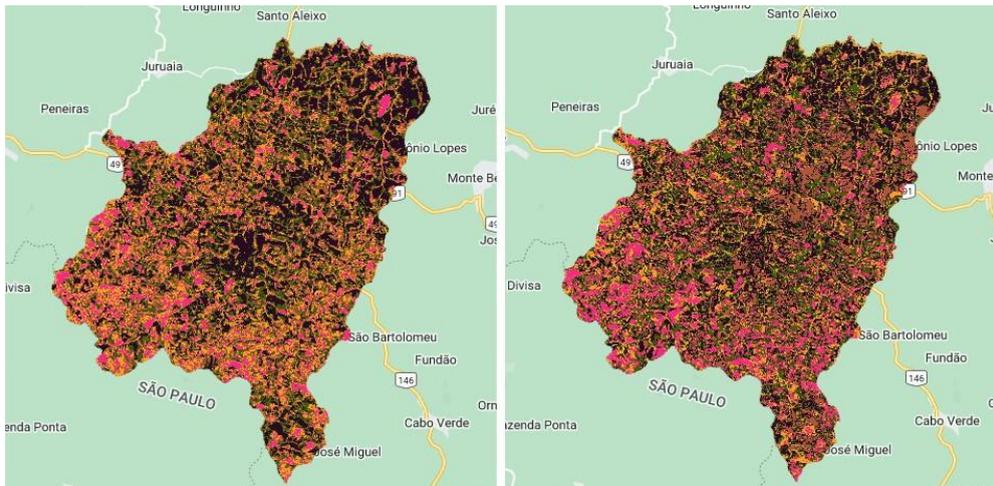
⁸ Disponível em: <<https://geemap.org/>>. Último acesso em 30/06/2023

⁹ Disponível em: <<https://pypi.org/project/PyCRS/>>. Último acesso em 30/06/2023

¹⁰ Disponível em: <<https://pypi.org/project/pysnp/>>. Último acesso em 30/06/2023

¹¹ Disponível em: <<https://developers.google.com/earth-engine/apidocs/ee-clusterer-wekakmeans>>. Último acesso em 30/06/2023

Figura 1 - Clusterização por K-MEANS RGB e K-MEANS NDVI



5. CONCLUSÃO

Neste artigo, foi explorada a aplicação do algoritmo K-Means na segmentação de imagens de satélite, concentrando na identificação de seleções em clusters. Utilizando as imagens do Sentinel 2, foi demonstrado os benefícios e as implicações práticas dessa técnica no campo das imagens de satélite. Os resultados mostraram que a utilização do índice NDVI para agrupamento produziu os melhores resultados, permitindo identificar com maior precisão segmentos de vegetação, asfalto e áreas descobertas, conforme indicado pela análise comparativa dos algoritmos na figura 1. Essas descobertas têm implicações práticas em diversas áreas, como monitoramento ambiental, detecção de mudanças na vegetação e avaliação de impactos ambientais. A combinação de inteligência artificial e imagens de satélite abre oportunidades para a análise e interpretação de dados geoespaciais, oferecendo insights valiosos para tomadas de decisão e estudos científicos.

REFERÊNCIAS

FITZ, Paulo Roberto; VIEIRA, Jeferson Cordeiro; SOARES, Mirlla Casimiro. O uso de polígonos de amostragem em classificações supervisionadas de imagens de satélite. *ENTRE-LUGAR*, v. 10, n. 19, p. 319-341, 2019.

GIRÃO, Inês; VIANA, Cláudia M.; ROCHA, Jorge. Tratamento de Dados Open source para Classificação de Imagens de Satélite. In: *Atas XII Congresso da Geografia Portuguesa, Geografias de Transição para a Sustentabilidade*. Universidade do Minho, 2020. p. 479-484.

KUMAR, Gautam et al. Performance of k-means based satellite image clustering in RGB and HSV color space. In: *2016 International Conference on Recent Trends in Information Technology (ICRTIT)*. IEEE, 2016. p. 1-5.

USMAN, Babawuro et al. Satellite imagery land cover classification using k-means clustering algorithm computer vision for environmental information extraction. *Elixir International Journal of Computer Science and Engineering*, v. 63, p. 18671-18675, 2013.